REPORT RESUMES

ED 003 362 24

EFFECTS OF AN INTENSIVE INSERVICE PROGRAM ON TEACHER'S CLASSROOM BEHAVIOR AND PUPIL READING ACHIEVEMENT.

BY- LOREE, M. RAY
ALABAMA UNIVERSITY, UNIVERSITY
REPORT NUMBER CRP-2608 PUB DATE 65
EDRS PRICE MF-$0.18 HC-$5.08 127P.

DESCRIPTORS- *PROBLEM SOLVING, *LEARNING PROCESSES, *CRITICAL THINKING, *PSYCHOEDUCATIONAL PROCESSES, *ELEMENTARY SCHOOL STUDENTS, JUNIOR HIGH SCHOOLS, EVALUATION TECHNIQUES, INTERVIEWS, UNIVERSITY, ALABAMA

THIS STUDY WAS THE COMPARISON OF THE INTELLECTUAL PROCESSES OF CHILDREN IDENTIFIED AS SUCCESSFUL PROBLEM SOLVERS WITH THE PROCESSES OF CHILDREN IDENTIFIED AS UNSUCCESSFUL. SPECIFIC ATTENTION WAS GIVEN TO THE PROCEDURE USED BY CHILDREN TO ARRIVE AT THE ANSWERS OF TEST QUESTIONS. THE PROCEDURE WAS THE INDIVIDUAL INTERVIEW TECHNIQUE FROM WHICH PROTOCOLS WERE OBTAINED. THE PROTOCOLS WERE ANALYZED FOR THE PURPOSE OF GENERATING HYPOTHESES FOR FUTURE TESTING. FOUR MAJOR CATEGORIES EMERGED FROM THE ANALYSIS OF THE PROTOCOLS AND MUST BE CONSIDERED IN RESEARCH ON PROBLEM-SOLVING ABILITY. THEY ARE (1) POSSESSION OF INFORMATION, (2) RECALL OF KNOWLEDGE, (3) EXTRACTION OF INFORMATION FROM THE PROBLEMS SITUATION, AND (4) COMBINING OPERATIONS (SUCH PROCESSES AS CLASSIFYING, ABSTRACTING, ANALYZING, SYNTHESIZING, AND EVALUATING). (JK)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

PROBLEM-SOLVING TECHNIQUES OF CHILDREN IN GRADES FOUR THROUGH NINE

Cooperative Research Project No. 2608

M. Ray Loree

University of Alabama
University, Alabama
1965

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

ED 003362

TABLE OF CONTENTS

CHAPTER PAGE

I. PROBLEM ON WHICH RESEARCH WAS FOCUSED 1

Objectives and Hypotheses 2

II. RELATED RESEARCH 5

Neobehaviorism 5
The Cognitive Position 7
Computer Model 9
Piaget's Developmental Stages 10
Factorial Approach 11
Logical Analysis 11

III. PROCEDURES 13

Phase 1 Selection of Subjects 13
Phase 2 Selection of Interview Test Items 15
Phase 3 Sample A Interviews 16
Phase 4 Sample B Interviews 17
Phase 5 Sample C Interviews 19
Phase 6 Scoring Procedures 19

IV. ANALYSIS OF THE DATA AND FINDINGS 22

Sample A 22
Hypotheses 28
Sample B 29
Sample C 38

V. SUMMARY AND CONCLUSIONS 59

Bibliography 66

Appendix I . . . . 20 Pages

Appendix II . . . . 32 Pages

LIST OF TABLES

TABLE PAGE

1. DISTRIBUTION OF CALIFORNIA SHORT FORM TEST OF MENTAL MATURITY I.Q. SCORES FOR SUCCESSFUL AND UNSUCCESSFUL SUBJECTS 15

2. SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE A SUBJECTS 27

3. COEFFICIENTS (rho) OF INTER-RATER CORRELATION ON ALL ITEMS, ALL CATEGORIES, SAMPLE B 30

4. SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE B SUBJECTS ON TEN COMMON ITEMS 31

5. COMPARISON OF MEAN RATING, BY CATEGORY, ON SELECTED ITEMS MISSED INITIALLY BY SUCCESSFUL AND UNSUCCESSFUL SAMPLES B & C SUBJECTS 33

6. SAMPLE B KRUSKAL-WALLIS ONE-WAY ANALYSIS OF VARIANCE, BETWEEN GRADES (Ten Common Items). 37

7. SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE B SUBJECTS ON FIVE ITEMS FOR GRADES 7, 8, AND 9 39

8. SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE B SUBJECTS ON FIVE ITEMS FOR GRADES 7, 8, AND 9 40

9. COEFFICIENTS (rho) OF INTER-RATER CORRELATION ON ALL ITEMS, ALL CATEGORIES, SAMPLE C 41

10. SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE C SUBJECTS ON TEN COMMON ITEMS 42

11. AVERAGE NUMBER OF TEST ITEMS "NOT UNDERSTOOD" BY SUCCESSFUL AND UNSUCCESSFUL SUBJECTS, (BY GRADES, SAMPLE C) 43

12. SAMPLE C KRUSKAL-WALLIS ONE-WAY ANALYSIS OF VARIANCE BETWEEN GRADES (TEN COMMON ITEMS) 47

13. SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE C SUBJECTS FOR GRADES 4, 5, AND 6 ON FIVE NON-COMMON ITEMS . . 49

14. SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE C SUBJECTS FOR GRADES 7, 8, AND 9 ON FIVE NON-COMMON ITEMS . . 50

15. PATTERNS OF WEAKNESSES OF SAMPLE B UNSUCCESSFUL SUBJECTS . . . 53

16. PATTERNS OF WEAKNESSES OF SAMPLE C UNSUCCESSFUL SUBJECTS . . . 53

ACKNOWLEDGEMENTS

I wish to thank Mr. W. W. Bristow, Principal of Eastwood Junior High School, Mr. Gordon Crawford, Principal of Stafford Elementary School, Mrs. Bess Hamner, Principal of Alberta Elementary School, Mr. Archie Hitson, Principal of Skyland Elementary School, Mr. Frank Kendall, Principal of Tuscaloosa Junior High School, and Mrs. Esther Westover, Principal of West End Elementary School, for making available the time and space needed for interviewing children. The assistance rendered by these principals and staff of their schools was a major contribution to the completion of this project.

Kenneth M. Parsley, Jr., collaborated with me for the first twelve months of the project. The development of the project proposal was a joint effort of Dr. Parsley and myself. During the period June 1, 1964 to June 1, 1965, Dr. Parsley assumed major responsibility for supervising the collection of data and attending to administrative arrangements for this study. I wish to acknowledge Dr. Parsley's fine contributions and regret that his departure from the University to assume a new position in Washington, D. C. prevented him from continuing with the project.

I wish to thank Educational Testing Service, Princeton, N. J. for permission to reproduce in this report items from the Sequential Test of Educational Progress.

What success this project has had is due in large measure to the intelligence, energy, and creativity of a group of graduate assistants who served as Research Assistants. Interviews were conducted by Frances Gunnels, Frank Lee, James Reaves, and Charles West. Each contributed many ideas and each assumed major responsibilities for conducting the analysis of the data reported in Chapter 4. The above group was joined by Joe Slate, Shirley Edwards, John Earls, and Robert Mallory in scoring the protocols and by Ronald Jones who helped in the analysis of the data.

Last but not least, I wish to thank Mrs. Susan Veal who served as secretary for the first nine months of the study, and to Mrs. Betty Jean Chard who typed reams of manuscript with skill and diligence during the final three months of the study.

M. Ray Loree

University of Alabama

Chapter I

Problem on which Research was Focused

The major problem of this research is to compare the problem-solving processes of children identified as successful problem solvers with those of children identified as unsuccessful problem solvers. The problem-solving processes of children in grades four through nine on one type of school-like problem-solving task--performance on standardized achievement test items--are investigated.

Attention is directed to the process through which children arrive at answers to test items. In each grade, successful and unsuccessful problem solvers were identified on the basis of their scores on an achievement test that required problem-solving ability. Processes employed by successful and unsuccessful problem solvers are compared, as well as the problem-solving processes of children at different grade levels.

Much research on problem solving, both with animals and humans, has been conducted by psychologists. However, as Travers (1963) points out, our knowledge of problem solving consists, more of a complex of information of "odds and ends," than of a cohesive body of knowledge. There are a number of possible reasons for this. First, a variety of tasks has been used in problem-solving research. A conceptualization of the problem-solving process that seems to fit well for one type of problem does not seem to be adequate for other types of problems. Second, problem-solving research has been conducted by workers with quite varied orientations. Stimulus-response psychologists, gestalt psychologists, information theorists, developmental psychologists, students of language structure, and educationists, all have contributed to the literature on problem solving. Thus, the terms "problem" and "problem solving" have been used by psychologists to describe a wide range of phenomena. No single description or theory of problem solving adequately takes into account all the research findings on the many types of problem tasks investigated in the psychological laboratories, nor the diverse questions that have been raised about the problem-solving process.

Fairly good agreement exists as to what constitutes a problem. A problem may be thought of as a gap between an initial state of affairs and a desired state of affairs. The process of reducing the gap involves more than just a recall of information. Before overtly responding, the problem solver must cognitively manipulate or structure information in the process of selecting or inventing a strategy designed to eliminate the gap. The process of reducing the gap between an initial state of affairs and a desired state of affairs may be different for one type of problem task than for another. In this research the problem tasks are achievement test items that call for more than rote memory.

Psychologists have differed in the nature of inferences they have made in observing the problem-solving process. Only overt responses can be observed as an individual solves a problem. The mediating processes that precede the overt responses are not directly observable. Research on problem solving has varied with respect to the kind of inferences made about mediating processes from observations of overt responses. Two classical viewpoints on problem solving emerged early in this century. Thorndike's observations of animals on problem-solving tasks led him to state the law of effect as the basic mechanism in learning. Thorndike (1911) postulated a trial-and-error process as the means through which an animal learns to solve a problem. Gestalt psychologists inferred, from their observations, foresightful planned behavior in solving problems, and introduced the notion of "insight." Much of the experimentation conducted during the first half of this century was oriented toward; first, investigating; and, later, reconciling these two classical viewpoints on problem solving.

Today, the number of different positions on the nature of problem solving has increased considerably. Yet the influence of Thorndike and the gestalt psychologists is still evident. The focus of many behaviorists' conception of problem solving is that of a sequence of events or a train of thought that includes stimuli resulting from previous internal responses and leading to another internal response. The inheritors of the Thorndikian position attempt to break down problem solving into elements and examine the sequence of those elements. The concept of structure is usually the focus of the cognitive theorists' explanation of problem solving. Miller, Galanter, and Pribram (1960), for example, contend that behavior is hierarchically ordered;with molar units, called "strategies," controlling molecular units, called "tactics." This appears to be more akin to the earlier gestalt position.

While the influence of earlier research on problem solving is evident in current research, there is at present an increased interest in what may be termed mediating cognitive processes. This is exemplified in research on information theories, in research on neurophysiological processes, in studies of language structure, in the return of interest to Piaget's life-long research on the thought processes of the child, in the investigations on verbal behavior, and in Rokeach's (1960) analysis of the open and closed mind. In addition, there has been a longer sustained interest of psychologists on thinking, concept formation, and creativity. In short, there is now a wealth of conceptualizations concerning thinking processes. However, relatively little of this research has been conducted with school-like tasks.

Objectives and Hypotheses

The major objectives of this research are to identify categories, constructs, and dimensions that are useful in describing (a) how the problem-solving processes of successful problem solvers differ from those of unsuccessful problem solvers; and (b) how the problem-solving processes of children in higher grades differ from those of children in lower school grades.

The kind of problems investigated in the study is restricted to certain social studies and science achievement test items found in standardized achievement tests. The major criterion in selecting these test items is that they measure more cognitively complex processes than the mere recall of information. This restriction in the type of problems investigated limits the generality of the usefulness of the categories, constructs, and dimensions identified in this study. Similarly, the term "successful problem solvers" in this study refers only to subjects who performed relatively well on an achievement test containing test items of a problem-solving type.

The term category in this report refers to a classification of the cognitive operations involved in the problem-solving process. For example, "defining the problem" may represent a category, within the complete problem-solving process, that is distinct from the category "gathering data needed to solve the problem."

The term construct in this report refers to an abstraction that characterizes an essential feature of cognitive operations. Thus, the terms "search model" and "set" represent constructs as do many of the terms used by Piaget (e.g., "intuitive thought," "concrete operations," "formal thought").

The term dimension in this report is used in two ways. First, it is used to refer to an ordering of categories along a continuum. Thus, the six categories of the cognitive domain of the Taxonomy of Educational Objectives (Bloom, 1956) are ordered along a "complexity" dimension. Categories within Guilford's (1959) theoretical model of the human intellect are ordered under three dimensions--"contents," "products," and "operations." The term dimension also is used in this report to mean simply a measure within a single category. Thus, "number of questions answered correctly" may constitute the dimension used in assessments of "information possessed." In assessing responses categorized as "creative," either "uniqueness of responses produced" or "number of creative responses produced" may constitute the dimension measured.

Two strategies are used in this study in searching for categories, constructs, and dimensions of use in describing qualitative and quantitative differences in problem-solving processes. The first strategy was to search the literature on problem solving for suggestions. Chapter II reviews research studies on problem solving that contributed directly or tangentially to this research. A second strategy was to present, in individual interviews, a sample of successful and unsuccessful problem solvers of different grade levels with a set of social studies achievement test items, and have each child "think aloud" as he attempts each item. Then, a comparison of the "thinking aloud" protocols is made in order to discover how the successful problem solvers differed from the unsuccessful problem solvers, and how children in higher grade levels differed from children in lower grade levels in their problem-solving processes. Chapter III describes in detail the procedures followed in pursuing this second strategy. Chapter IV contains an analysis of the kinds of differences in problem-solving processes found in this first sample of successful and unsuccessful problem solvers.

The initial sample of children to be interviewed in this study is referred to throughout this report as Sample A. The question arises as to whether or not the categories, constructs, and dimensions found to be of value in describing differences in problem-solving processes for children in Sample A are also of value in describing differences found in other samples of children, and for other samples of achievement test items. To answer this question, two additional samples of children were interviewed.

Students in a second sample of successful and unsuccessful problem solvers--Sample B--were interviewed, using the identical achievement test items used for Sample A.

A third sample of successful and unsuccessful problem solvers--Sample C--was drawn. These students were interviewed using science achievement test items, rather than social studies items.

In summary, the major objectives of this study are:

1. to explore the utility of the individual interview technique as a means of identifying and understanding the problem-solving processes of children as they work upon achievement test items of a problem-solving type;
2. to analyze the protocols of the testing interviews in order to find out what categories, constructs, and dimensions are useful in describing the differences between the problem-solving processes of the successful and the unsuccessful problem solver, and the differences among the problem-solving processes of children of different grade levels;
3. to test whether the categories, constructs and dimensions of problem-solving processes found to be discriminating within one sample of successful and unsuccessful problem solvers also is found to be discriminating for a second sample of successful and unsuccessful problem solvers, using the same test materials; and for a third sample of subjects, using test materials from a different subject-matter area.

The analysis of the protocols of the testing interviews for the children in Sample A was designed to generate hypotheses concerning the differences in the problem-solving processes between the successful and the unsuccessful problem solvers, and among children at different grade levels. These hypotheses are stated in Chapter IV.

Chapter II

RELATED RESEARCH

In order to solve a problem, an individual must close a gap between an initial state of affairs and a desired state of affairs. Before overtly responding to a problem situation, the problem solver may cognitively manipulate or structure information in the process of selecting or inventing a strategy designed to eliminate the gap. Research by psychologists on problem-solving processes has been extensive. Johnson (1955) and Vinacke (1952) report a multitude of research studies. In recent years, psychologists have manifested an increased interest in what may be termed mediating cognitive processes. Two books of readings (Harper, Anderson, Christensen, and Hunka, 1964; Anderson and Ausubel, 1965) provide a cross-section of research reports on cognitive processes.

While there is a wealth of conceptualizations concerning thinking processes to be found in the psychological literature, relatively little research has been concerned with testing the various conceptualizations by using school-like tasks. One strategy planned for this study was to examine the applicability of various conceptualizations on problem-solving processes to an analysis of the cognitive processes used by children on achievement test items of a problem-solving type. It is relevant, therefore, to examine briefly some of the major theoretical orientations to problem solving. Six orientations will be discussed. These are: (1) neobehaviorism; (2) cognitive position; (3) computer model; (4) Piaget's developmental approach; (5) factorial approach; and (6) logical analysis. Particular attention will be paid to the research methods employed, the variety of tasks used, and the major categories, constructs, and dimensions generated from the various studies conducted.

Both the neobehaviorism and the cognitive positions stem from psychological theories with long histories. No attempt will be made to trace the early contributions to these positions. The concepts advanced by only two or three of the many representative researchers of each orientation will be discussed.

Neobehaviorism

A stimulus-response paradigm can explain well reflex behavior, but does not explain thinking processes. To account for thinking, it is necessary to introduce some form of mediating cognitive process. Stimulus-response psychologists who are interested in cognition have postulated that the external stimulus evokes an internal response, which produces an internal cue that is connected to the overt response. It is assumed that the internal stimulus and response events follow the same principles that operate in observable stimulus-response relationships (Osgood, 1953; Goss, 1961; Kendler and Kendler, 1962; Staats and Staats, 1963).

Objective experimentation constitutes the major method of analysis of neobehaviorists in their study of mediating cognitive processes. For example, Judson, Cofer, and Gelfand (1956) performed an experiment in which they used Maier's two-string problem as the task.* The subjects in the experiment learned lists of words in serial order before being presented with the two-string problem. For the experimental group of subjects, the words rope--swing--pendulum were inserted into the word list in that order. Other groups learned these same words, but not in that order. It was hypothesized that the prior verbal learning of the sequence rope--swing--pendulum would mediate the behavior of making a pendulum and swinging it, thus solving the problem. The results of the experiment tended to support the hypothesis.

Staats and Staats (1963) identify three behavioral mechanisms involved in problem solving: tacting, verbal response chains, and overt motor responses mediated by the individual's own verbal responses. Tacting is a process through which a verbal response is elicited in the presence of stimulus objects (e.g., the verbal response rope in Maier's two-string problem).

More generally, tacting is a term used by neobehaviorists to describe a conditioning process through which meaningful responses become attached to stimulus objects. Verbal response chains involve the sequencing of verbal responses set off by a problem situation. Reasoning, for the neobehaviorist, is this sequence of verbal responses, overtly or covertly expressed, which leads to the solution of the problem. Thus, the string stimulus evokes the verbal response rope, which in turn evokes the verbal responses swing and pendulum. Finally, problem-solving behavior is elicited by the word-response sequence (e.g., the behaviors of tying a weighted object on one piece of cord and swinging it to form a pendulum are elicited by the word-response sequence rope--swing--pendulum).

Various tasks have been used by neobehaviorists in their research on cognitive processes. Numerous studies involve problems of the puzzle variety such as the two-string problem. In recent years, studies have used a wide variety of verbal-learning material in investigating verbal associations. The central orientation of these studies is that of determining the sequence of events within the problem-solving process.

From the educator's point of view, one shortcoming of many of the experiments conducted by neobehaviorists on problem solving, is that the tasks provided in the experiments have little apparent resemblance to the kind of problem-solving tasks that a child encounters in school. Usually, reports of experiments contain no information which provides clues as to the degree to which the problem-solving ability acquired during an experiment might be expected to transfer to school-like tasks. However, research that is oriented toward neobehaviorism can be structured so as to be relevant to the kinds of problem solving that take place in schools. Anderson (1964), for example, conducted an experiment in which "the goal was to teach six- and seven-year old children a classical method of experimental science, the strategy of varying each factor in succession while holding other factors constant." Concept attainment problems, similar to those employed by Bruner, Goodnow, and Austin (1956), were used as instructional materials and were organized into a sequential small-step teaching program. Some evidence was obtained of transfer to concept-attainment problems, other than those contained in the training program. Protocols in Anderson's study were scored for certain characteristics in the sequence of instances selected by each child. These characteristics were the number of unnecessary selections, the number of invalid conclusions and the number of solutions reached.

---

*In Maier's two-string problem, the subject is given the task of tying two strings, hanging from the ceiling, which were too far apart to reach by hand. The solution required tying a weighted object to one string, thus making a pendulum of it. By swinging the pendulum, the subject is enabled to grasp the swinging pendulum while holding the other string.

The research described in this report differs from the kind of experimental studies conducted by neobehaviorists. Interviews were not conducted in order to establish verbal associations which could be used later in solving the problem presented. Nevertheless, the interpretation of the neobehaviorists of the problem-solving process did suggest that we be alert to the verbal associations elicited by successful and unsuccessful problem solvers by the achievement test items presented.

The Cognitive Position

Problem solving, for the cognitive theorist, is much more of an active effort on the part of the problem solver to search for the structure of relationships among the elements of the problem situation, to seek to determine the constraints of the problem--constraints that eventually lead to the formation of hypotheses. Bruner (1961) attributed major importance to the acts of discovery for the problem-solving process, and he suggests that school children would improve their problem-solving abilities if they were afforded more experiences in learning through discovery. One type of problem used by Bruner in his research has been the game of Twenty Questions. For example, a child is told that a car has gone off the road and hit a tree. The child is instructed to ask questions that can be answered "yes" or "no", in order to discover the cause of the accident. Problem-solving strategies of different children then are compared. Some children ask questions designed to narrow the range of possible hypotheses (e.g., "Was there anything wrong with the driver?"). In Bruner's words, these children employ "constraint location" as a technique before formulating hypotheses. Other children began their questioning with a specific hypothesis (e.g., "Was the driver rushing to the doctor's office for an appointment and the car got out of control?"). For Bruner, learning to structure the information possessed by the problem solver is of prime importance in developing problem-solving abilities. To quote:

> "I would urge now in the spirit of an hypothesis that emphasis upon discovery in learning has precisely the effect upon the learner of leading him to be a constructionist, to organize what he is encountering in a manner not only designed to discover regularity and relatedness, but also to avoid the kind of information drift that fails to keep account of the uses to which information might have to be put" (Bruner, 1961, p. 25).

It may be noted that in the Twenty Question variety of problem, success in solving the problem is contingent upon the problem solver's receiving more specific information than is contained in the initial statement of the problem. The problem solver needs to ask himself the question, "What additional information do I need to help me solve the problem?" There is no doubt that many problems encountered in life are of this type. However, success in solving all problems is not highly contingent upon securing additional information. In some cases, all the information needed to solve the problem either, is contained in the statement of the problem or, it has been previously acquired by the problem solver. The major tasks of the problem solver appear to be that of extracting relevant information from the statement of the problem and/or retrieving information in a way that meets the demands of the problem. Maier's two-string problem is of this latter type. So too are many problem-solving type test items contained in achievement tests.

Over a period of years, Rimoldi and his co-workers (1962, 1964) have utilized problems which require, for their solution, more information than is contained in the initial statement of the problem. For example, the instructions for one problem used by Rimoldi is:

> "A college choral group is composed of freshmen, sophomores and juniors. The chorus has three voices or parts which are high, medium, and low. The questions and answers below give vital information concerning the group. From these facts you are to find the number of juniors singing the middle or medium part" (1962, Appendix I, p. 25).

For a problem such as the above, Rimoldi typically presents a subject with a set of cards containing questions that he may ask in order to reach a solution. Some of the questions listed are useful in solving the problem posed, while others are not. The subject is free to choose any card, in any order he wishes. The answer to each question is printed on the reverse side of the card. Some of the seventeen questions and answers provided for the above problem were:

	Questions		Answers
1.	How many juniors are in this college?	1.	1567
3.	How many sophomores are in the middle voice?	3.	10
4.	How many chorus members are there?	4.	76
5.	How many girls are in the chorus?	5.	45
6.	How many sophomores are in the chorus?	6.	28
10.	How many low voices are there?	10.	28
14.	How many chorus members sing the high voice?	14.	24
16.	How many freshmen sing the middle voice?	16.	6

For the above problem, the number of medium voices in the chorus can be obtained when the answers to questions 4, 10, and 14 are known. The number of juniors singing in the medium part (8) can then be deduced from the answers to questions 3 and 16. Rimoldi has developed a number of objective measures that reflect the efficiency of the problem solver in seeking and using information. These include the number of cards required to reach a solution, the number of correct solutions, and the utility or efficiency of the sequence of questions in arriving at solutions to the problem.

A modification of Rimoldi's technique was utilized in this research project in our study of the processes through which students solved achievement test items.

The quality of the problem-solving strategies of students may be effected by the mode of representation of a problem, as well as by its difficulty level. Bruner (1964) describes three modes of representation: enactive, utilizing motor responses; iconic, utilizing images of perceptual events; and symbolic, involving the manipulation of abstractions. In the present study, test items of varying difficulty level were selected, and (for some items) the interview procedure was such as to shift the mode of representation from the symbolic to the iconic, and (for one item on the science test) from the iconic to the enactive.

Cognitive theorists emphasize the importance of "structuring" in the problem-solving process. However, there is not agreement as to the role of structuring. Gagne (1962, 1965) focuses attention on the learner's possession of subordinate problem-solving abilities which provide learning sets that are prerequisite to and enter into the process of solving a problem. Ausubel (1963) focuses attention on the role of structure in reception learning. For Ausubel, structure constitutes the ideational orientation of the learner as he encounters new learning materials.

Computer Model

Methods of solving problems that have been built into modern electronic computers closely resemble those used by humans solving similar problems (Newell, Shaw, and Simon, 1958; Hovland, 1960). Newell, Shaw and Simon had college students "think aloud" as they worked on problems in symbolic logic, and then compared their methods with detailed records of the processes employed by a computer program.* The resemblances of the problem-solving processes of their computer program to human problem-solving are summarized by the writers as follows:

> First, and perhaps most important, it is in fact capable of finding proofs for theorems--hence incorporates a system of processes that is sufficient for a problem-solving mechanism. Second, its ability to solve a particular problem depends on the sequence in which problems are presented to it in much the same way that a human subject's behavior depends on this sequence. Third, its behavior exhibits both preparatory and directional set. Fourth, it exhibits insight both in the sense of vicarious trial and error leading to "sudden" problem solution, and in the sense of employing heuristics to keep the total trial and error within reasonable bounds. Fifth, it employs simple concepts to classify the expressions with which it deals. Sixth, its program exhibits a complex organized hierarchy of problems and subproblems (p. 161).

---

*Newell, Shaw and Simon point out that "thinking aloud" is behavior, (not introspection) just as truly as circling the correct answer on a paper-and-pencil test is behavior. The inferences made concerning the cognitive processes that mediate the "thinking aloud" behavior, of course, are not behavior.

The information processed by a computer must be translated into symbols that the computer program can handle. Thus, the first step in solving a problem is one of abstraction. Symbols are used to represent objects (descriptions of the given, desired, and various intermediate possible situations), differences between pairs of objects, and operations that are capable of bringing about changes in the object to which they are applied. Simon and Newell (1962) point out that the program must be arranged to attain three types of goals: (1) transformation goals: to transform object a (the given situation) into object b (the desired situation); (2) difference reducing goals: to eliminate or reduce difference d between objects a and b; and (3) operation application goals: to apply operation q to object a.

Miller, Galanter, and Pribram's (1960) description of information-processing strategies is somewhat similar to that of Newell, Shaw and Simon. Miller, Galanter, and Pribram use the term "image" to represent the goal to be attained, and the term "plan" for the strategy used to reach the goal. In executing a plan, a constant feedback occurs in which operations performed are checked against the requirements of the image. For Miller, Galanter and Pribram, a plan consists of four stages: (1) a test to ascertain differences between the present and the desired state of affairs; (2) an operation to reduce these differences; (3) a second test to reevaluate differences; and (4) an exit when the desired state of affairs has been attained. The whole sequence is termed a TOTE unit (test, operate, test, exit). Solving a single problem may include subordinate TOTE units within the total plan.

Some modification of a computer model may have some value in examining the processes through which children reach solutions to problem-type achievement test items. The memory storage unit represents the past learning of the child. Unlike the computer, the child may encounter difficulty in retrieving information. The test item itself may contain additional information, as well as posing the question. Information processing will involve generating strategies in which past learning and information contained in the test item are combined to meet the requirements of the question posed.

Piaget's Developmental Stages

Hunt (1961) and Berlyne (1957, 1960) have noted the similarity between Piaget's conceptions of mediating processes and those developed by researchers working on computer information processing. Hunt (1961) states "Piaget's observations of the homely interactions of the child with his everyday environment demonstrate empirically the formation of a vertical hierarchy of operations for the processing of information to guide action" (p. 109). Piaget (1947) contends that the structure of intelligence undergoes a continuous transformation as a result of interaction between maturational and experiential forces. Three of Piaget's stages of development--intuitive thought (ages 1½ or 2 to 4 years), concrete operations (7 or 8 to 11 or 12 years), and formal operations (11 or 12 to 15 or 16 years)--are of concern to this study. Many (but not all) investigators who gathered data designed to test Piaget's classification of developmental levels have obtained results that are in general agreement with Piaget (Peel, 1959; Lovell, 1961; Case and Collinson, 1962). Case and Collinson's data reveal a marked tendency for older children to use more mature thinking processes, but that all children in his sample of children (aged 7-17) answered at least one question at the formal thought level, and that children tended to regress to less mature levels when presented with complex materials. In investigations conducted at the University of Alabama preliminary to this research study, much more difficulty was experienced in applying Piaget's criteria for stages of development to the "thinking aloud" protocols of children on achievement test items in social studies than in science. Piaget's criteria for level of operations worked best for test items in which children were called upon to supply an explanation for a phenomenon.

Factorial Approach

Guilford's (1959) conceptualization of the "structure of intellect" may be considered to be a fifth type of approach to the study of cognitive mediating processes (Ausubel, 1965, pp. 14-16). Guilford's factorial studies have led him to postulate an organization of intellectual abilities into a three-dimensional matrix formed by five classes of operations (cognition, memory, divergent thinking, convergent thinking, and evaluation); four kinds of content (figural, symbolic, semantic, behavioral); and six types of products (units, classes, relations, systems, transformations, and implications). In all, the matrix yields a total of 120 abilities or cells. Guilford's distinction between divergent and convergent thinking has been of particular value in research on creativity.

Logical Analysis

The analysis of problem-solving that probably has influenced educators most is that presented by Dewey (1910). Dewey analyzed "the complete act of thought" into the following stages:

1. Becoming aware of the problem
2. Clarifying and defining the problem
3. Searching for facts and formulating hypotheses
4. Evaluating proposed solutions
5. Experimental verification

For certain types of problems, Dewey's analysis does not appear to be adequate for studying differences in problem-solving processes. Bloom and Broder (1950) in a study (similar in many respects to this research study) had six academically successful and six academically unsuccessful college students "think aloud" as they attempted to solve examination test questions. The data, so obtained, did not fit well into Dewey's five stages. Bloom and Broder summarized the differences in problem-solving processes as revealed in their protocols under four headings:

1. Understanding the nature of the problem
2. Understanding the ideas contained in the problem
3. General approach to the solution of the problem
4. Attitude toward the solution of problems

The Taxonomy of Educational Objectives, Cognitive Domain (Bloom, 1956) represents another type of analysis of cognitive functioning. Statements of the educational objectives of schools and colleges were examined and the objectives classified into six categories: knowledge, comprehension, application, analysis, synthesis, and evaluation. Like Guilford's model, the taxonomy attempts to identify a comprehensive set of abilities, rather than a description of cognitive functioning. The difference may be more apparent than real, for, as Ausubel (1965) points out, "an intellectual ability, after all, is nothing more or less than a functional manipulation of a distinct and identifiable cognitive process as expressed in a range of individual performance or capacity differences" (pp. 14-15).

The present research study is designed to explore differences in the problem-solving processes of successful and unsuccessful problem solvers in grades four through nine. Previously conducted research on problem solving and analyses of the problem-solving process served as a starting point in searching for constructs, categories, and dimensions that might differentiate the successful from the unsuccessful problem solver. The studies reviewed above constituted the major sources of ideas for this research project.

Chapter III

PROCEDURES

The procedures for this study fall into six phases:

Phase 1 involved procedures for forming three samples (A, B, and C) of ten successful and ten unsuccessful problem solvers for each grade level, four through nine.

Phase 2 involved the selection of test items to be included in the interviews of the children of each sample.

Phase 3 included the development of an Interview Guide for the children of Sample A, the interviewing of each child in Sample A, and the transcription of each interview recording.

Phase 4 included the revision of the Interview Guide for use with the children in Sample B, the interviewing of each child in Sample B, and the transcription of each interview recording.

Phase 5 included adapting the Interview Guide to the questions selected from science tests, the interviewing of each child in Sample C, and the transcription of each interview recording.

Phase 6 included the development of scoring procedures, the application of these procedures to some 350 protocols, and the analysis of the data so obtained.

Phase 1

Selection of Subjects

The procedures that seem appropriate to be used to identify successful and unsuccessful problem solvers depend, in part, upon one's view of the nature of problem solving. A variety of measures of problem-solving abilities would seem appropriate, if problem solving is viewed as a single process or a generalized ability brought into play in any problem-solving task. If, on the other hand, one views problem-solving processes as varying markedly, according to the kind of task involved; then, the measure of problem-solving ability should be restricted to a measure of performance on the kinds of problem-solving tasks that one is investigating. The diverse descriptions of problem-solving processes contained in the research literature lend credence to this second view of problem solving. For this reason, a single measure of problem-solving ability was employed in this study.

Approximately 120 children in each grade level, four through nine, were administered either, Form 4A (grades 4, 5, 6) or, Form 3A (grades 7, 8, 9) of Social Studies, Sequential Test of Educational Progress. These 720 children were drawn from intact classrooms of public schools in Tuscaloosa, Alabama. The test items included in STEP were reviewed in order to ascertain whether or not some items should be eliminated from the scoring because they appeared to tap only rote memory. A decision was reached to accept the total score on the test as a measure of the kind of problem-solving ability which this study is exploring. For each grade, the forty children with the highest scores (the top third) were identified as "successful" problem solvers; the forty children with the lowest scores (the bottom third) were identified as "unsuccessful" problem solvers. The investigators then randomly drew, from each grade level, the names of ten children from the top third and ten children from the bottom third. These 120 children became the subjects for Sample A, to be interviewed, using achievement test items drawn from Form 3B and 4B of the Social Studies, Sequential Test of Educational Progress. In a similar fashion, the names of an additional 120 children were randomly drawn to become the subjects for Sample B (to be interviewed using the same social studies test items). The names of sixty children from Sample A and an additional sixty children (who were not included in either Samples A or B) were drawn to become the subjects for Sample C (to be interviewed using test items from the Science STEP test). Thus, the total number of planned interviews was 360 (ten successful plus ten unsuccessful problem solvers (20), times six grade levels (6), times three groups or samples).

From the protocols obtained from the subjects in Sample A and from suggestions contained in the research literature, it was hoped to develop categories and dimensions that would be useful in describing differences between problem-solving processes of successful and unsuccessful problem solvers, and to develop a scoring procedure to quantify those differences. The categories and dimensions identified in Sample A then were applied to a fresh sample of subjects--Sample B. The generality of the usefulness of the categories generated from Sample A was tested on a third Sample--Sample C--with achievement items drawn from new subject-matter--science.

The focus of this investigation is upon problem-solving processes, rather than upon status characteristics that appear to influence problem-solving processes. The only two status characteristics involved in the design of this study are: (1) successful problem solvers versus unsuccessful problem solvers; and (2) grade level. However, information on a third status characteristic--intelligence test performance--was obtained. The appropriate form of the California Test of Mental Maturity, Short Form, was administered to the children of all three samples for grades four through six. For grades seven and nine, the school records included test results on the California Test of Mental Maturity, Short Form. This data was not used, however, to equate groups of successful and unsuccessful problem solvers.

Table 1 contains data that indicates that successful subjects were more able (as measured by the California Test of Mental Maturity) than the unsuccessful subjects. However, some successful subjects appear to have less general ability than some unsuccessful subjects. General mental ability appears to facilitate the kind of problem solving required by the Social Studies STEP test, although above average ability apparently does not guarantee successful performance.

Table 1

DISTRIBUTION OF CALIFORNIA SHORT FORM TEST OF MENTAL MATURITY I.Q. SCORES FOR SUCCESSFUL AND UNSUCCESSFUL SUBJECTS

	SUCCESSFUL				UNSUCCESSFUL			
GRADE	RANGE	Q_1	Q_2	Q_3	RANGE	Q_1	Q_2	Q_3
4	95-135	108	111	115	65-116	91.5	103	106
5	87-136	100	108.5	116	62-115	81	87	93
6	85-139	115	120	123.5	60-113	78	88.5	103
7	101-131	111	118	123	59-114	86	97	102
8*								
9	106-123	109	119	123	68-118	82	99.5	110

* Data for grade 8 was not available.

Phase 2

Selection of Interview Test Items

The test items to be administered within the interviews of subjects were selected from Forms 3B and 4B of the STEP test. Items were selected from the social studies tests for Samples A and B, and from the science tests for Sample C. Appendix I, pages 2 and 5, contains the ten social studies test items selected to be administered to the Sample A and B subjects, in all grades; and Appendix II, pages 2, 4, 6, 9 and 12, contains the ten science test items selected to be administered to the Sample C subjects, in all grades.

The following criteria were considered in selecting these items:

(1) The items should demand more than mere recall of information;
(2) The items should contain content familiar to grades four, five, and six children;
(3) The difficulty level of the items should be appropriate for children in all grades, four through nine.

Social studies and science curricula for grades four through nine were examined in the process of applying the above criteria. Item difficulty was determined by consulting item analysis data contained in the Teacher's Manual for the STEP tests.

In addition to the ten items selected for all grades, five items for grades four, five, and six subjects (See Appendix I, pp. 11-12, and Appendix II, pp. 15, 16, 20) and five items for grades seven, eight and nine subjects (See Appendix I, p. 15, and Appendix II, pp. 23, 25, 27, 30) were selected.

Phase 3

Sample A Interviews

The Interview Guide to be used for Sample A emerged from a series of staff meetings attended by the two principal investigators and the four graduate research assistants who were to conduct the interviews. Previous studies at the University of Alabama on problem-solving processes (Roberts, 1964, Possien, 1964, Peake, 1964) seemed to indicate that gaps would occur in the information derived from problem-solving protocols unless the interviewer probed, systematically, specific hypotheses concerning the problem-solving process. For example, to investigate whether or not successful problem solvers possessed more background information relevant to the problem posed than did unsuccessful problem solvers, it is necessary to make certain that data be obtained on the amount of background information possessed by each child for each problem included in the interviews. At the same time, the systematic coverage of a series of questions, generated by preconceived notions concerning the nature of the problem-solving process, reduced the possibility of discovering new categories, constructs, and dimensions of the problem-solving process. In order to give children an opportunity to reveal unanticipated processes and still systematically explore certain anticipated processes, the Interview Guide was divided into two parts--a non-probing portion, which was largely unstructured, and a probing portion, which was structured, to explore a number of potential dimensions that might serve to discriminate the successful from the unsuccessful problem solver.

The interview procedures adopted for the children in all three samples consisted of non-probing and probing portions. Preceding the formal interview, each subject was given practice test items, so he could become accustomed to "thinking aloud" while attempting to answer a test question.

Two other features of the interview procedures should be mentioned. These are: (1) the use of flaps covering segments of certain test items; and (2) the use of card sorts on certain test items.

Use of flaps

A neobehavioristic approach to the study of problem solving entails investigating the nature of the covert stimulus-response associations that intervene between the appearance of an observable stimulus and the behaver's overt response (e.g., Osgood, 1953; Judson, Cofer, and Gelfand, 1956; Goss, 1961; Staats and Staats, 1963). Problem solving is thus viewed as a sequence of covert stimulus-response associations. The interview procedures developed in this study attempted to gain information on the sequence of steps followed by subjects through the use of flaps which were attached to segments of test questions. The subject was required to request the interviewer to raise a particular flap, whenever he wished to direct his attention to a particular segment of the test question. For example, in the series of questions based upon an imaginary country (See Appendix I, p. 5), a flap was placed over the map and each one of the six questions in the series. The interviewer noted the sequence as the subject looked from the question to the map, back to the question, and so forth. The interviewer also tried to ascertain what information the subject sought each time he refocused his attention.

Use of card sorts

A modification of the Rimoldi card sort technique (Rimoldi, et.al., 1962, 1964) was used as a means of testing a subject's comprehension of the ideas contained in a problem. Cards containing either information needed to solve a given test item or information irrelevant to its solution were presented to the subject. The subject then was required to select the card or cards containing information needed to answer a given test item, and to tell why the information on the cards selected was needed in order to answer the test question. (See Appendix I, pp. 6, 7, 16, 17).

The interviews

Ten successful and ten unsuccessful problem solvers in each grade (four through nine) were included in Sample A. Interviews were conducted in rooms assigned by the principals of five cooperating schools. Each interview lasted approximately one hour. The four graduate research assistants who conducted the interviews met periodically with the two principal investigators to discuss revisions in the interview procedures. An interviewer was not informed as to whether a subject to be interviewed was classified as a successful or an unsuccessful problem solver. However, the interviewer usually did form some opinion as to the problem-solving ability of the subject during the course of an interview.

Recordings of the interviews were transcribed during the same period as the interviews were conducted. For Sample A, some interview recordings were found to be unusable because of their poor sound quality. This difficulty was eliminated in subsequent samples.

Phase 4

Sample B Interviews

At the completion of the interviewing of Sample A subjects, the Interview Guide was revised in preparation for the Sample B interviews. Parts B, of Appendix I, contain the revised Interview Guide. The differences in the problem-solving processes, evidenced in the Sample A protocols, seemed to be classifiable into four major categories:

(1) possession of background information;
(2) retrieval of information;
(3) extraction of information contained in statement of a problem;
(4) combining operations.

The probing portions of the interviews for both Samples B and C subjects were directed towards gaining data which would be used in assessing each subject on each of the four above categories.

An analysis of Sample A protocols revealed a number of instances where inability to retrieve information appeared to be the source of a child's difficulty in solving a problem. The child appeared to lack required information when first presented with a problem. Later, the series of questions concerning the problem appeared to stimulate the recall of needed information. Therefore, "retrieval of information" was singled out as a category separate from that of "possession of information." Certain questions were presented, initially, in a free response form, so that the child had minimal recall cues. Additional recall cues were provided later, if needed.

The "extracting information" category included an understanding of the problem posed, and the ability of the child to direct attention to key words in the statement of a problem or to key portions of a diagram or chart accompanying a problem. "Searching behavior" constituted an important subcategory of "extracting information", particularly for test items containing maps or diagrams. The ability to differentiate the relevant from the irrelevant appeared to be a possible measurement dimension of this subcategory. West (1965) found this selection variable contributed to the difficulty level of test items. Performance of subjects on card sorts was taken, on some items, as evidence of the subject's ability to extract information.

The "combining operations" category appears to be the most complex of the four categories. It is possible, through a logical analysis, to hypothesize what information a child must possess and retrieve, and what information must be extracted from the statement of the problem in order to answer the test item. These elements then must be combined through some operation (e.g., classifying, comparing, logical reasoning, structuring, applying, analyzing, synthesizing, evaluating, TOTE units, hypotheses formation, and hypotheses testing, and so forth). Logically identifying the elements to be combined within achievement test items proved to be a less difficult task, than assigning a "tag" name to the cognitive processes involved in an operation. The operations involved in the protocol materials on some items in Sample A can be described equally well in terms of "structuring" or "analyzing;" TOTE units or hypotheses formation and testing. The operations involved for some items appeared to be quite different from the operations involved for other items. In short, the protocol materials for Sample A did not suggest a simple classification of the processes involved in combining operations. Hence, a systematic exploration of a "process" dimension of combining operations is not reflected in the Interview Guide used for Samples A or B. Some of the protocol material for Sample A suggests that combining operations could be analyzed along a molar-molecular dimension. To illustrate, some subjects in Sample A seemed to be formulating and testing hypotheses in the process of responding to certain items. But this process appeared to be of a much more molecular level, than the process followed by the scientist in thinking about a scientific problem. The range of problem tasks included in this study was not conducive, however, to an exploration of a molar-molecular dimension of combining operations. In summary, the Interview Guide for the combining operations category that emerged following the Sample A interviews focuses attention on the elements to be combined. The elements to be combined may be derived from a retrieved information source, or may need to be extracted from a posed problem. Once the elements become visible to a subject, the number of elements to be combined may be a factor in the difficulty of a problem.

The actual interviewing of the subjects in Sample B proceeded as with the subjects in Sample A. The same graduate students conducted the interviews. The use of a revised Interview Guide constituted the only change in procedure.

Phase 5

Sample C Interviews

The procedures followed in adapting the Interview Guide to science test items and the methods of conducting the interviews paralleled the procedures followed with the Sample B interviews. The test items selected, first, were analyzed in terms of the four major categories: possession of information; retrieval of information; extracting information; and combining operations. Interview questions next were designed to provide a basis for assessment of each child with respect to each category. The non-probing portion of the interview, and certain questions in the probing portion, provided opportunities for the child to exhibit unanticipated problem-solving strategies.

Some minor changes in interview strategies were introduced for the subjects in Sample C. It had become increasingly evident during the interviewing of Sample B subjects that many children did not understand the problem posed in a test item and consequently failed the item. Therefore, each Sample C subject was asked to state, in his own words, what problem was posed in each test item. A second change in interview procedures involved the use of more visual aids within the interview. This was done in order to explore whether an inability to visualize a problem might be a contributing factor to failure on a test item. A third change in interview procedures stemmed from the fact that the science test items emphasized understanding of principles, the ability to supply explanations for phenomena, and an understanding of the scientific method, more than did the social studies test items. Questions designed to determine whether or not a subject could think of, or recognize, additional examples of a scientific principle were injected into the interview. Also, a number of questions were asked to determine a subject's understanding of the scientific method.

The actual interviewing of the subjects in Sample C proceeded essentially in the same way as with subjects in Samples A and B, with the Interview Guide adapted to science test items constituting the only change in procedure.

Phase 6

Scoring Procedures

Scoring procedures for the protocols of all the samples involved a preliminary identification, for each test item, of the information a subject would need to have and retrieve, the information a subject would need to extract from the test item, and the combining operations through which the subject verbalized relationships between elements of a problem. Following an initial identification of the elements subsumed under each category for each item, weights or ratings from 1 to 5 were assigned to various combinations. Thus, for a particular item in which two bits of information were needed to answer the question, a rating of five was assigned to the subject who demonstrated possession of both bits of information; a rating of three was assigned to the subject who lacked one bit; and a rating of one was assigned to the subject lacking both bits of information. Two independent graders rated each protocol; then discussed their independent ratings and reached a consensus rating for each item on each protocol. Unforseen difficulties in scoring were discussed in staff meetings, and the scoring procedure guide was clarified.

The C portions of Appendix I, and Appendix II, list the elements subsumed under each of the four categories, for each item, in the social studies and science tests.

Test item Number 29, from Part I, Form 4B, of the science STEP test will serve to illustrate the scoring procedure.

The item reads:

The sun is the only object in our solar system that makes and gives off large amounts of heat and light. This means that we see the moon because it

A reflects light from the sun
B gives off its own light
C is not in our solar system
D reflects much light from other planets

In the individual interviews, the above item was presented to ten successful and ten unsuccessful problem solvers of each grade, four through nine. In the non-probing portion of the interview each child was encouraged to "think aloud" as he attempted to answer the item. The following series of questions was then posed to each child in the probing portion of the interview (See pp. 7 and 8, Appendix II):

1. Can you tell me what the problem asks you to do?
2. Why did you select the answer you did?
3. (a) What is a solar system?
   (b) What is our solar system?
   (c) Is the moon in our solar system?
4. What is the differences between a planet and a star?
5. Is the sun a planet or a star?
6. Do planets give off large amounts of heat and light?
7. Does the earth give off its own light or reflected light? Why or why not?
8. Does the moon give off its own light or reflected light?
9. Where does the light in our solar system come from?
10. Why did you not select answer (--), (--), or (--)?

Scoring on "possession of information" on this item was based upon a subject's responses to the above interview questions 3, 4, and 5. A rating of 5 was assigned, if the subject answered correctly each of the three questions; a rating of 1 was assigned for incorrect answers to all three questions; and intermediate ratings of 4, 3, 2 were assigned for partial knowledge.

Scoring on "retrieval of information" was based upon a subject's answer to interview question 2. A subject received a high rating for recalling information about our solar system or about planets and stars in the process of justifying his answer to the test question. For all social studies and science test items the "retrieval of information" score was based upon recall of information, either in the non-probing portion of the interview, or in explaining his answer to a test item. "Possession of information", on the other hand, might be evidenced at any point within the interview.

Scoring for "extracting information" for Question 29 was based on: (1) a subject's understanding of the problem posed (interview question 1); and (2) whether or not the subject evidenced comprehension of the statement, "The sun is the only object in our solar system that makes and gives off large amounts of heat and light," in responding to interview question 6, 7, 8, and 9, or in responding to interview question 2.

Scoring on "combining operations" was based upon a subject's response to either interview question 2 or 10. The highest rating was awarded, when the subject combined recalled information and information contained in the test item in answering interview question 2. An intermediate rating was awarded when, after the series of interview questions 3-9, the subject clarified his thinking on the test item and was able to answer, satisfactorily, interview question 10.

One problem arose in developing scoring procedures for the science test. It seemed reasonable to assume that all of the children possessed all the information required to cope with certain test items. The problem posed in such test items involved, primarily, extracting information from data provided in a test item and combining that information in a logical manner in the process of arriving at an answer. For example, in the series of test items 20, 22, 24, 25 based on Form 4B, Part II, of the science STEP test (See Appendix II, p. 12), indoor and outdoor hourly readings are presented on a temperature chart, and pictures of an indoor and an outdoor thermometer are presented. The test items based upon the chart and the thermometer seem to involve, primarily, an ability to read the temperature chart and the thermometers, and an ability to correlate information extracted from the two sources. For items of this kind, a rating of 3 was arbitrarily assigned to "possession of information" and "retrieval of information," unless a subject displayed unusual and unanticipated possession of or defect in information. Seven science test items (and one social studies item) were of this type.

Chapter IV

ANALYSES OF THE DATA AND FINDINGS

Sample A

At the completion of the Sample A interviews, the four graduate students who conducted the interviews were asked to list ways in which the successful problem solver differed from the unsuccessful problem solver. Sample A protocols were studied by the chief investigators with the same purpose in mind. From these analyses of Sample A data and from a consideration of the categories, constructs, and dimensions discussed in the research literature, four major categories emerged. These categories, which seemed to subsume more specific observed differences, were:

1. possession of information
2. retrieval of information
3. extracting information
4. combining operations.

The first category represents a status characteristic of the problem solver, whereas the remaining three categories represent process characteristics of the problem solver.

1. Possession of information

The information that seemed helpful to a subject in coping with a problem included:

a. an understanding of concepts such as "natural environment," "latitude," "symbol," "leisure time," "slope of land," which appear in the statement of test items;
b. understanding of ideas contained in test items, such as "lowest elevation above sea level," and "proportion of time;"
c. more general information derived from reading or experience and related to the specific test item--such as the processes involved in manufacturing shoes, the reason why people engage in different activities, and population trends in the United States;
d. specific facts pertinent to test items, such as "rivers flow toward the sea," and "the United States lies north of the equator;"
e. information that constitutes a part of a skill, such as map reading or interpretation of charts and diagrams.

The category, "possession of information," corresponds, in part, to Bloom and Broder's category, "understanding the ideas contained in the problem."

2. Retrieval of information

The successful problem solver, on reading a test item, evidenced recall of information pertinent to the problem posed. The unsuccessful problem solver appeared to lack required information when initially presented with a test item, but later in the interview sometimes demonstrated possession of the required information. The interesting question is whether the process of retrieving information is teachable. The orientation of the neobehaviorists may be particularly fruitful in delineating the conditions controlling processes such as "tacting" and "verbal response chains." Ausubel (1960) who has conducted research from a structuralist orientation, has demonstrated that "advanced organizers" facilitate the learning and retention of meaningful verbal material. Protocol materials in Sample A indicate that "talking around" a test item stimulates the recall of pertinent information for both the successful and unsuccessful problem solver.

3. Extracting information

The successful problem solver in Sample A appeared to be superior to the unsuccessful problem solver in:

a. grasping the problem posed by a test item;
b. searching for pertinent information and discarding irrelevant facts within a test item;
c. focusing attention on key elements (facts, concepts, ideas, phrases) within a test item;
d. selecting from memory storage information pertinent to the test item;
e. profiting from experience on one test item in working on subsequent test items.

The category, "extracting information," subsumes a number of categories and constructs to be found in the literature on problem solving, including Bloom and Broder's category, "general approach to the solution of the problem;" Dewey's categories, "clarifying and defining the problem" and "searching for facts and formulating hypotheses;" the first T (testing) of the TOTE unit; and the constructs "structure," "set," and "search model."

Inadequate reading ability appeared to be a major factor explaining some of the unsuccessful problem solvers' difficulty in understanding the problem posed in a test item, particularly in the lower grades. The interviewers found it necessary to read test items to some children. The successful problem solvers read directions and questions carefully, quickly chose some point at which to attack the problem, were able to pick out key words, were able to decide what knowledge was relevant to the problems, and discarded irrelevant facts.

The successful problem solver's efforts to understand what question was being asked in a test item often could be described as an effort to discover structure inherent within the test item. The following extracts of protocol materials are illustrative of differences between successful and unsuccessful sixth graders in structuring test materials. The test items involved ordering pictures so as to tell the story of shoes. Question 21 asks which picture should come first. Question 22 asks which picture should come after the tannery (See Appendix I, pp. 11 and 12).

Successful sixth-grade subject

Int: Here are some questions and here are some pictures, what would you like to do first?
S: I'll look at the pictures first.

Int: All right, I'll open this flap and let you look at the pictures. What are you looking at in the pictures?
S: Well, I'm looking at the, uh, the boxes and pictures and things, shoe store, stockyard, tannery, shoe factory, pasture, shoes in use. Now I'm reading the title of it (Reads). Now I'll study it and see which one will come first...Now I'll want to see the first question.
Int: I'll open this up and let you see Question 21.
S: I'm reading the title (lead to test item) first. (Reads question and alternative responses.) Uh D, the pasture would come first.
Int: All right.
S: I want to look at Question 22.
Int: O.K., I'll open this up and let you see Question 22.
S: I'm reading the title first. (Reads question and responses.) O.K....I'm looking at the title again. (Reads question over.) The shoe factory,F, the shoe factory.

(The following dialogue occurred later in the probing portion of the interview.)

Int: For Question 21, you said D, pasture,was the answer, why did you think so?
S: Well, the cattle would have to eat to get their hides good and fat.
Int: All right, are you sure of that answer, you don't want to change it?
S: No ma'am.
Int: Why did you select F, the shoe factory,as the answer to Question 22?
S: Well, after the tannery, they have to go to the shoe factory to be put in boxes....
Int: What else is there in a shoe factory besides packaging?
S: Well, men work there and, uh, I guess they put heels on the shoes...and put soles and things on.

Unsuccessful sixth-grade subject

Int: Let's turn over to the first set of questions. Here are some pictures and some questions. What would you like to do first?
S: I'd like to do the pictures first.
Int: Why?
S: I think it might help me.
Int: Remember to tell me everything you see and everything you think.
S: I see shoes...and some cows and some trees. (long pause)
Int: What would you like to do now?
S: See the questions. (Reads Question 21) Number C, the shoe factory.
Int: What would you like to do next?
S: See Question 22. (Reads the question. Interviewer helps her with "tannery.") G--the pasture.

(The following dialogue occurred later in the probing portion of the interview.)

Int: For Question 21 you said the shoe factory was the correct answer.

Why did you think so? (pause) Do you think the shoe factory should come first?

S: No ma'am.

Int: You want to change your answer? What would you want to change it to? (pause)

S: The stockyard.

Int: Let's look over here at these pictures. Now what are these pictures the story of? (long pause)

Int: Look all over the page and see if you can find anything that might tell you what these pictures are about.

S: Shoes. (Interviewer has subject read out loud the heading to the pictures. "These pictures tell the story of shoes. They are not in the right order.")

Int: They are not in the right order. Now what does that mean to you? (long pause) What do you think you have to do with them if they are not in the right order?

S: Put them in the right order.

After a tremendous amount of work (three pages of protocol material), this sixth grader finally was able to order the pictures correctly. The word "tannery" caused particular difficulty. This subject seemed unable to work constructively on the sequence problem until she understood the specific meaning of each individual picture.

Possibly the search for information of the problem solver is guided by a "search model," "sets," and "hypotheses," but the nature of the guides that directed searching behavior on the part of children was difficult to infer from Sample A protocol materials.

4. Combining operations

The "combining operations" category subsumes all processes--classifying, comparing, abstracting, analyzing, synthesizing, evaluating, structuring and so forth--through which more than one element of a problem is cognitively manipulated. An element of a problem may be a fact, concept, principle, idea or theory. Elements may be imbedded in the statement of a problem, or may be contained in memory storage. Whereas memory is the basic process underlying retrieval of information; and perception is basic to extracting information; thought is basic to combining operations.

To generate a classification of all possible combining operations would require, as a basis, a large sample of problem-solving tasks. The combining operations exhibited in the protocols of Sample A were classified as follows:

a. searches for information contained in the statement of the problem in the light of the requirements of the problem;
b. relates recalled information to the requirements of the problem;
c. relates recalled information to information contained in the statement of the problem.

Selecting one of the four reasons given to explain "Why do people act as they do?" (given information) as the best fit for the question, "Why do Americans go to ball games?" (requirement of the problem)--is an example of the first type of combining operation. The processes in this particular example may correspond to a TOTE unit, thus:

Test: to find why many Americans go to ball games
Operate: to look for best answer among the four given alternatives
Test: to apply test--"Is this the reason?" to each alternative
Exit: to end search when satisfied that best fit has been located.

Ordering the pictures in the set of questions based upon "The story of shoes," is an example of the second type of combining operation. The problem solver needs to draw upon his knowledge about the sequence of events leading up to the use of shoes (recalled information) and relate this knowledge to a question posed in a test item (requirement of the problem).

Relating recalled information to information contained in the statement of the problem, a third type of combining operation, is illustrated in protocol material on Question 14, Social Studies STEP, Form 3B, Part I. This test item, based upon a map of an imaginary country, asks what direction you would have to travel to reach the United States. The following example of logical reasoning was found:

a. United States is north of the equator (retrieved information).
b. The map of the imaginary country is south of the equator (extracted information).
c. Therefore, one must travel north from the imaginary country to reach the United States (logical deduction).

Motivation, attitudes toward the test situation, confidence in solutions reached, and persistence in the face of experienced difficulty were reported by the interviewers as factors affecting problem-solving performance. However, for a number of reasons, these non-cognitive factors were not included as a separate scoring category. First, these factors were more apparent and easier to evaluate in a live interview situation than in a typescript of an interview. Second, while some evidence of these non-cognitive factors can be found in some protocols, in many other protocols the evidence is either missing or equivocal. Third, non-cognitive factors appear to operate in quite complex and diverse ways.

The C portions of Appendix I detail the scoring procedures followed in generating scores on each of the four major categories. Table 2 contains a summary of the scores obtained by successful and unsuccessful Sample A subjects.

Table 2

SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE A SUBJECTS

	SUCCESSFUL		UNSUCCESSFUL		
	Median	Range	Median	Range	Significance level of differences*
Category 1					
Possession of information					
Grade 4	50	34-64	39.0	23-47	.01
Grade 5	55	45-63	35.0	17-45	.001
Grade 6	66	53-71	41.0	25-55	.001
Grade 7	65.5	56-74	53.0	29-65	.01
Grade 8	65.5	54-74	51.5	38-64	.01
Grade 9	63.5	59-70	49.0	33-66	.01
Category 2					
Retrieval of information					
Grade 4	46	29-68	34	21-42	.01
Grade 5	51	40-69	30	17-41	.001
Grade 6	60.5	55-70	36	19-52	.001
Grade 7	61.5	50-73	46	23-62	.01
Grade 8	59.0	50-71	45	25-56	.01
Grade 9	61.0	46-71	46	15-60	.01
Category 3					
Extracting information					
Grade 4	46	32-66	35	23-48	.05
Grade 5	53.5	42-65	27	15-43	.001
Grade 6	63.0	56-69	36	16-49	.001
Grade 7	68.5	55-74	49.0	26-64	.001
Grade 8	62.0	53-74	51.5	35-61	.01
Grade 9	70.0	51-73	51.0	30-60	.001
Category 4					
Combining operations					
Grade 4	42.0	28-73	28.0	19-38	.01
Grade 5	47.5	37-62	26.0	15-33	.001
Grade 6	60.5	50-71	34.0	15-42	.001
Grade 7	60.0	48-69	45.0	20-60	.01
Grade 8	53.0	42-73	38.0	25-56	.05
Grade 9	59.0	47-67	44.5	15-56	.001
Category 5					
No. of correct answers					
Grade 4	8.0	4-12	4.5	2-8	.01
Grade 5	8.0	7-11	5.0	3-8	.001
Grade 6	12.0	9-14	3.0	3-10	.001
Grade 7	10.0	5-13	5.0	3-9	.01
Grade 8	9.0	5-14	5.0	2-8	.01
Grade 9	10.0	6-13	6.0	3-12	.01

*Mann-Whitney U tests.

The table lists the median score for each category and the median number of questions answered correctly by successful and by unsuccessful subjects for each grade. Ranges in scores are also presented. Mann-Whitney U tests were used to test the probability that the distributions of scores for successful and unsuccessful grade groups were drawn from the same population. Significance levels of the differences are reported in Table 2. Scores reported in Table 2 are based upon fifteen items. Hence, the total possible score in each category is 75 and the total possible number of questions that could be answered correctly in the interview is 15. However, while ten test items were common for all grades, five of the items presented to grades seven, eight, and nine were different and more difficult than five of the items presented to grades four, five, and six.

The data reported in Table 2 indicates that successful and unsuccessful problem solvers do differ in measures in each category. The evidence of differences in category scores and in number of correct answers among children of different grades is not as clear as the differences between successful and unsuccessful problem solvers.

Some overlap was found in the score distributions between successful and unsuccessful subjects in twenty-three of the twenty-four grade-category comparisons contained in Table 2. However, in all six grades, there is an overlap in the distributions of the "number of correct answers" scores. Some children who were classified as unsuccessful problem solvers performed better in the interview situation than did some children who were classified as successful problem solvers. There are a number of possible reasons for this. Possibly some children were classified incorrectly. Some children who were classified as unsuccessful problem solvers may have performed well below their usual level on the initial STEP test used to assess problem-solving ability. It is also possible that the interview constituted quite a different type of test situation than the pencil-and-paper test. For some students, the interview may have been a supporting situation that elicited their best performance. For other students, the interview situation may have lowered performance level.

Hypotheses

Three major objectives of this study are stated in Chapter I. The first objective concerns the utility of the interview technique as a means of identifying and understanding the problem-solving processes of children as they work upon achievement test items of a problem-solving type. This objective will be discussed later in this chapter. The second objective entailed an analysis of the Sample A protocols in order to see what categories, constructs, and dimensions differentiate between the problem-solving processes of successful and unsuccessful problem solvers of different grade levels. Four categories were found to differentiate. These categories subsume or are closely related to a number of categories and constructs that previously have been used to describe the problem-solving process. The third objective may be stated in the form of a hypothesis as follows:

1. Children, in grades four through nine, classified as successful problem solvers are superior to children, of comparable grade level, classified as unsuccessful problem solvers in measures

of (a) possession of information, (b) retrieval of information, (c) extracting information, and (d) combining operations when, first, social studies achievement test items are used in interviews as a basis for obtaining measures and when, second, science achievement test items are used as the basis for measures.

The procedures followed in testing the above hypothesis limits, in a number of ways, the generality of any conclusions reached. Success in problem solving is judged solely upon the performance of pupils on a social studies achievement test of a problem-solving type. It is not known whether the pupils who are classified as successful or unsuccessful problem solvers would be classified similarly for their performances on other types of problem-solving tasks. Similarly, the measures on the four major categories are based upon the responses of selected children to selected achievement test items in social studies and science.

The analyses of the Sample A protocols suggested further problems which deserve study. These problems may be stated in the form of hypotheses as follows:

2. Children classified as successful problem solvers in the upper grade levels are superior to children classified as successful problem solvers in the lower grade levels in measures of (a) possession of information, (b) retrieval of information, (c) extracting information, and (d) combining operations when, first, social studies, and, second, science achievement test items are used in interviews as a basis for obtaining measures.

3. Children classified as unsuccessful problem solvers in the upper grade levels are superior to children classified as unsuccessful problem solvers in the lower grade levels in measures of (a) possession of information, (b) retrieval of information, (c) extracting information, and (d) combining operations when, first, social studies, and, second, science achievement test items are used in interviews as a basis for obtaining measures.

Sample B

Rater reliability

The Spearman rank correlation coefficient was used to estimate the correlation between the rating assigned in each grade, for each of the four major categories, and for a total score on process, by two independent raters. As Table 3 indicates, twenty-one of the thirty correlations were .80 or above. Rater agreement was less correlated in the fourth and the eighth grades. The coefficients reported in Table 3 represent conservative estimates of the reliability of consensus ratings, because large discrepancies in a rating usually occurred as a result of one rater overlooking portions of a protocol. The consensus rating corrected such errors. Also, the pooling of judgments from two observers yields increased reliability in a manner comparable to that found in the doubling of the length of a test (p. 466, Guilford, 1965).

Table 3

COEFFICIENTS (rho) OF INTER-RATER CORRELATION
ON ALL ITEMS, ALL CATEGORIES, SAMPLE B

		Category				
Grade	N	1 Possession	2 Retrieval	3 Extraction	4 Combining	5 Total
4	20	.409	.611	.517	.800	.696
5	20	.847	.860	.848	.949	.802
6	20	.844	.793	.590	.902	.941
7	20	.863	.871	.915	.941	.930
8	20	.853	.794	.735	.693	.848
9	20	.865	.886	.869	.863	.972

Differences between the successful and unsuccessful

Hypothesis 1 involves a comparison of scores, on each of the four categories, of successful and unsuccessful problem solvers of comparable grade levels. Table 4 contains a summary of the scores obtained by successful and unsuccessful Sample B subjects on the ten common test items. The table lists the median score for each category and the median number of questions answered correctly by successful and unsuccessful subjects for each grade. Again, as in Sample A, overlap was found in the score distributions between successful and unsuccessful subjects in all but one of the twenty-four grade-category comparisons. Also, again, there is an overlap in the distribution of scores in all six grades on the number of test items answered correctly.

Hypothesis 1, stated in the null form (Ho) is: There is no difference between children classified as successful problem solvers, and children of comparable grade level, classified as unsuccessful problem solvers in measures of (a) possession of information, (b) retrieval of information, (c) extracting information, and (d) combining operations, when social studies achievement test items are used in interviews as a basis for obtaining measures. The Mann-Whitney U test was used to test this hypothesis. The level of confidence was set at .05 ($a = .05$). Since there was a basis for prediction of direction, a one-tailed test was used. Table 4 reports the significance levels of differences for the twenty-four grade-category comparisons of successful and unsuccessful groups of subjects. Mann-Whitney U tests are also reported for differences, at each grade level, for the number of correct answers given in the interviews to the test items. Except for the five grade four comparisons and one comparison in category 3 at the grade six level, all differences were found to be significant.

The fact that differences in all four categories and in total scores between successful and unsuccessful grade four subjects were not significant is somewhat puzzling. The interviewers did experience the greatest difficulty in eliciting verbal responses from pupils at the grade four level.

Table 4

SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE B SUBJECTS ON TEN COMMON ITEMS

	SUCCESSFUL		UNSUCCESSFUL		
	Median	Range	Median	Range	Significance level of differences*
Category 1 Possession of information					
Grade 4	30.5	25-36	27.0	21-35	n.s.
Grade 5	34.0	27-44	28.0	10-39	.01
Grade 6	43.0	31-48	34.0	25-44	.05
Grade 7	48.5	40-50	35.5	24-42	.001
Grade 8	43.0	37-48	34.0	28-45	.001
Grade 9	46.0	36-50	37.5	31-45	.001
Category 2 Retrieval of information					
Grade 4	28.0	22-32	24.5	16-38	n.s.
Grade 5	31.5	25-44	23.5	10-43	.025
Grade 6	39.0	25-45	33.0	22-41	.025
Grade 7	41.5	35-49	29.5	17-34	.001
Grade 8	43.5	34-49	32.5	29-42	.01
Grade 9	44.0	35-49	39.5	24-38	.001
Category 3 Extracting information					
Grade 4	31.5	22-44	29.5	27-34	n.s.
Grade 5	36.0	32-44	29.0	10-44	.001
Grade 6	40.5	29-46	36.5	29-43	n.s.
Grade 7	43.5	34-49	33.5	22-44	.001
Grade 8	43.5	38-48	38.0	26-43	.001
Grade 9	44.5	36-50	39.0	28-45	.002
Category 4 Combining operations					
Grade 4	29.5	14-35	25.0	13-31	n.s.
Grade 5	34.0	22-44	22.5	10-39	.001
Grade 6	38.0	23-44	28.5	21-36	.01
Grade 7	39.5	36-46	31.0	18-36	.001
Grade 8	39.5	30-47	33.0	23-43	.001
Grade 9	40.0	29-50	33.5	26-39	.01
Category 5 No. of correct answers					
Grade 4	4.0	1-8	4.0	1-5	n.s.
Grade 5	4.5	3-8	3.0	0-6	.01
Grade 6	6.0	5-7	4.0	1-8	.01
Grade 7	7.0	4-9	4.0	3-7	.001
Grade 8	6.5	3-9	6.0	1-8	.01
Grade 9	7.0	2-8	5.0	4-7	.05

*Mann-Whitney U tests based on all items.
n.s. = not significant at the .05 level.

This may be a factor contributing to lower inter-rater reliability found at the grade four level. An examination of the protocols of the two unsuccessful grade four problem solvers who obtained high scores on the four categories revealed that each had reading difficulty and that, while each child seemed to have initial difficulty in structuring a test item, each seemed to improve during the probing portions of the interview. It may be that the STEP Social Studies Test used for classification purposes measured reading ability as much as problem-solving ability for our Sample B subjects and hence resulted in a misclassification of some pupils.

A comparison was made of the performances of successful and unsuccessful problem solvers on selected test items which both initially answered incorrectly. Items were selected which were missed frequently by successful subjects. The successful problem solvers performed somewhat better on each of the four categories in almost every instance. The all-grade totals show a difference in favor of the successful problem solvers on all four categories. Table 5 contains the averege ratings of successful and unsuccessful problem solvers on a 5-point rating scale. Ratings are shown for all grades and an all-grade total is also presented. It will be noted that the average rating is low, even for the successful pupils. However, the successful pupils performed somewhat better than did the unsuccessful pupils, with differences in ratings ranging from .655 on Category 1 to .382 on Category 3.

On the basis of an examination of protocols of both successful and unsuccessful problem solvers, some generalizations can be made.

Successful problem solvers who initially answered the question incorrectly tended to change to the correct answer with little or no probing or help from the interviewer. The same was not true of the unsuccessful problem solver, who reached the correct answer, if at all, only after considerable probing and discussion with the interviewer. The following excerpts from protocols serve to illustrate this point. (See Item 8, *Appendix I, p. 2)*

Successful 8th grade subject

Int: Question 8, answer 4. Why?
S: I think now, number F, 2.
Int: Why?
S: Well..uh..your friends and neighbors would go to a baseball game and I think this one would go along with 2.

Unsuccessful 6th grade subject

Int: Question 8, answer 1, why?
S: 'Cause they have some kind of rules for baseball.
Int: Is there a law that says you have to go to baseball games?
S: No, sir.
Int: Will you reword the question for me?
S: (Reads Question)
Int: O.K., why do Americans like to attend baseball games? Now look at the four choices and tell me which one you think would be most nearly correct.
S: Can I change my answer?
Int: Yes--what would you like to change it to?

Table 5

COMPARISON OF MEAN RATING, BY CATEGORY, ON SELECTED ITEMS MISSED INITIALLY BY SUCCESSFUL AND UNSUCCESSFUL SUBJECTS IN SAMPLES B AND C

	SAMPLE B		SAMPLE C	
	Successful	Unsuccessful	Successful	Unsuccessful
Category 1				
Grade 4	2.80 (40)	2.38 (40)*	2.85 (20)	2.94 (33)
Grade 5	2.97 (34)	2.31 (42)	2.96 (27)	3.00 (38)
Grade 6	3.52 (31)	3.15 (34)	3.09 (21)	2.90 (37)
Grade 7	3.76 (37)	2.60 (58)	3.10 (21)	2.58 (31)
Grade 8	3.58 (41)	3.26 (45)	3.43 (14)	2.88 (25)
Grade 9	4.15 (41)	3.22 (51)	4.19 (16)	2.63 (24)
Total	3.474(224)	2.819(270)	3.21 (119)	2.80 (188)
Category 2				
Grade 4	2.33	1.95	2.77	2.50
Grade 5	2.41	1.98	2.40	2.61
Grade 6	2.61	2.47	2.62	2.43
Grade 7	2.84	2.02	3.10	2.52
Grade 8	3.27	3.00	3.00	3.24
Grade 9	3.56	2.57	2.88	2.75
Total	2.861	2.326	2.78	2.64
Category 3				
Grade 4	2.78	2.93	3.52	2.92
Grade 5	3.23	2.26	3.00	3.12
Grade 6	2.97	3.29	3.71	3.40
Grade 7	3.51	2.68	3.81	2.61
Grade 8	3.69	3.31	3.86	2.96
Grade 9	3.73	3.31	4.13	2.92
Total	3.334	2.952	3.64	3.01
Category 4				
Grade 4	2.18	1.95	2.33	1.76
Grade 5	2.44	1.83	1.85	1.61
Grade 6	2.68	2.24	2.19	2.16
Grade 7	2.86	2.09	2.43	1.77
Grade 8	3.07	2.35	3.29	1.80
Grade 9	3.00	2.39	2.88	2.33
Total	2.714	2.148	2.43	1.89

*Numbers in parentheses indicate the total number of incorrect responses on selected items by subjects who initially answered the item incorrectly.

S: To 3, because it would not be good for their health if they... Change it to 2..(reads 2)
Int: Why?
S: Because their families and neighbors go there and I guess they would like to go there too.

Even when both missed the question, the successful problem solver seemed to structure the problem more correctly than the unsuccessful problem solver. The successful problem solver, even when he answered the question incorrectly seemed to be better able to put elements of the problem into a sequence then the unsuccessful problem solver, as is illustrated by the following excerpts. (See Item 21, Appendix I, pp. 11, 12)

Successful 5th grade subject

Int: Question 21, answer tannery, why?
S: (pause)
Int: What does 21 ask?
S: Well..which..(reads question) well..I don't know what it means ...I know..but I just can't say it out in the right words...
Int: What are the pictures about...What do all these pictures tell?
S: Oh...it tells something about the story.
Int: The story of what?
S: The shoe story. It tells about the things you do.
Int: O.K., where do you start?
S: ...the shoe factory...they have to make them.
Int: Where do they get the material to make the shoes? What are the shoes made of?
S: Leather.
Int: And where does leather come from?
S: From cattle.
Int: And where does the story of cattle begin?
S: The pasture.
Int: Does that mean the pasture comes first?
S: Yes ma'am.
Int: After the pasture what comes next?
S: The tanner...(pause)...they take them to the stockyard.
Int: and then...
S: ...they take it to the tannery..the shoe factory...the shoe store.

Unsuccessful 5th grade subject

Int: Question 21, answer shoe factory, why?
S: ...cause it is the first picture and it says up there which picture should come first.
Int: What do you think the question is asking?
S: (reads question)
Int: O.K., let me give you a clue. The question is asking...in the story of making shoes which picture should come first. Now look at those pictures. If you were going to make a pair of shoes, what would be the first thing you would have to start with?
S: The shoe factory?
Int: Would you start with the shoe factory, what do they have at a shoe factory?
S: They have machines to make shoes.

Int: What do they make shoes from?
S: Leather.
Int: Where do they get the leather from?
S: The stockyard.
Int: I mean what do they get the leather from?
S: The tannery.

The successful problem solver who was initially unable to answer the question correctly frequently attempted to reason through a problem on the basis of faulty information, such as the meaning of a word. The unsuccessful problem solver did not appear to do this to the same extent. (See Item 8, Appendix I, p. 2)

Successful 8th grade subject
Int: Question 8, answer H, why?
S: Because some people have different environments than others.
Int: What does environment have to do with the baseball game?
S: The way people act around them...what they see...all in everyday life.
Int: Would you restate this question?
S: Baseball is attended by a great number of Americans...and it says why?...and I say it's H, Number H, because of the natural environment in which they live.

Unsuccessful 5th grade subject
Int: Question 8, answer H, why?
S. (reads question)
Int: What does H mean to you?
S: Cause like if you were sick and you went to a little league ball game it would help your health and....Hey, I meant number 3.
Int: What if you had a cold and went to a ball game and it rained...
S: Yeh...but most people bring along an umbrella when its going to rain.
Int: Do you think you might go to a ball game because you're required to go?
S: No ma'am.
Int: Do a lot of (people) go because their families, friends or neighbors are going?
S: No ma'am...they can...I think 3 is closest.

Successful 8th grade subject (See Item 14, Appendix I, p. 5)
Int: Question 14, answer west, why?
S: Well, the U.S. is in the western hemisphere and well, I just think this would be out maybe in the Atlantic Ocean...So you would have to travel west to reach the U.S.

Unsuccessful 8th grade subject
Int: Question 14, answer west, why?
S: West? uh...(pause)...it would be east cause I had my directions mixed up.
Int: Why east? Why would the U.S. be east of there?
S: I don't know.
Int: What information would you need to have about this country to make you know for sure?
S: The name of the country.

Int: How would that help you?
S: Maybe I could recognize the name of the country and on a map I could find the U.S.

The successful problem solver who missed the question seemed to be more likely than the unsuccessful problem solver to realize that additional information was needed to solve a problem. The following excerpts are illustrative. (See Appendix I, p. 5)

Successful 5th grade subject
Int: Question 14, answer south, why?
S: I don't know, I just guessed.
Int: What information could I give you that would help you answer this question?
S: Is the U.S. above or below the equator?
Int: The U.S. is above the equator. (pause) Is there any other information I can give you?
S: Is this country above or below the equator?
Int: This country is mostly below the equator.
S: Then you would travel...north.
Int: Why would you travel north?
S: Because this country is below the equator and the U.S. is above the equator and you'd have to go north to reach it.

Unsuccessful 5th grade subject
Int: Question 14, answer east, why?
S: Cause you would go east to get to it.
Int: Why?
S: You could have a map where it is....(long pause)..you'd have to travel east, wouldn't you?
Int: Would you? (pause) Is there any information I can give you?
S: (long pause) My answer is right isn't it?
Int: No. (long pause) Is there any information I can give you that would help you get the right answer?
S: (Names off directions)
Int: That's right...Is there anything I can tell you?
S: ...No sir...
Int: ...are you sure there's no information that you would like to have to solve this problem?
S: No.
Int: You're postive?
S: Yes sir.

Differences between grades

An inspection of Table 4 reveals that children in the upper grade levels appear to do somewhat better than children in the lower grade levels on each major category and on the number of items answered correctly. The Kruskal-Wallis One-Way Analysis of Variance was used to test whether the scores from the different grades could have been drawn from the same population (Hypotheses 2 and 3). It was found that the null hypothesis could be rejected at the .05 level or better for all categories, for both successful and unsuccessful problem solvers. Table 6 summarizes these findings.

Table 6

SAMPLE B KRUSKAL-WALLIS ONE-WAY ANALYSIS OF VARIANCE, BETWEEN GRADES (Ten Common Items)

	Level of Significance
Category 1 (Possession of information)	
Successful	.001
Unsuccessful	.001
Category 2 (Retrieval of information)	
Successful	.001
Unsuccessful	.01
Category 3 (Extracting information)	
Successful	.001
Unsuccessful	.001
Category 4 (Combining operations)	
Successful	.001
Unsuccessful	.01
Category 5 (Number of correct answers)	
Successful	.001
Unsuccessful	.05

It may be noted, from Table 4, that the medians do not increase by grade in a uniform manner. For the successful problem solvers, gains are rapid up to grade seven. The grades eight and nine pupils, in general, did very little better than the grade seven pupils. This may reflect very little improvement in problem-solving processes in the upper junior high school grades or the little improvement may be an artifact attributable to the level of difficulty of the test items included in the interviews. If all pupils in a grade obtained the highest possible scores, then the median would be 50. Some of the medians included in Table 4 were close to the maximum.

Mann-Whitney U tests were used to test the significance of differences between adjacent grades. Most of the differences were not significant. For the unsuccessful group, however, differences were significant at the .05 level or better between grades five and six.

An analysis was made of the scores obtained by successful and unsuccessful subjects on the five items included only in the grade seven, eight, and nine interviews. Table 7 includes a summary of scores obtained by successful and unsuccessful Sample B subjects.

Kruskal-Wallis One-Way Analysis of Variance was used to test the probability that the grades seven, eight and nine samples could have been drawn from the same population. Only on the "information possessed" category were differences found to be significant beyond the .05 level among the grade samples of successful problem solvers. Among the grade samples of unsuccessful problem solvers, significant differences beyond the .05 level were found on the first three categories. A considerable overlap in scores was found between grades for both successful and unsuccessful samples. Medians for the samples of successful problem solvers approach the maximum of 25. A comparison of Tables 4 and 7 suggests that care must be exercised in selecting test items of appropriate difficulty level in order to detect improvements in the problem-solving processes of pupils. In general, the more difficult item appears to be more sensitive to differences between grades.*

For comparison purposes, scores on the five items included only for grades four, five, and six were analyzed. Table 8 contains a summary of scores. The five test items provided an inadequate test ceiling for the successful problem solvers, although significant differences beyond the .05 level were found for the first two categories. A significant difference was found only on the "possession of information" category among score distributions for grade samples of unsuccessful problem solvers.

In general, hypotheses two and three concerning superior performances of higher grade level children over those of lower grade level children for each of the category scores are borne out by the data for both successful and unsuccessful problem solvers.

Sample C

Rater reliability

Spearman rank correlation coefficients between the independent ratings on each grade-category and on total score for each grade were computed for Sample C. Table 9 contains the inter-rater reliability estimates. In general, ratings were in closer agreement on the science protocols than on the protocols on materials in social studies. As errors in ratings were corrected by consensus ratings, the coefficients reported in Table 9 represent conservative estimates of inter-rater reliabilities.

---

*While selecting more difficult test items may result in a test that is more sensitive to improvements in students' performances from grade to grade, it also may result in lowering the discrimination between the performances of successful and unsuccessful students within grades (Loree, 1948, and pp. 495-497, 1965). To investigate whether or not this occurred, the ten common items were ranked in difficulty level according to the Teacher's Guide for the STEP test. It was found that unsuccessful problem solvers tended to do about as well as successful problem solvers on the more difficult test items. For example, 33 out of 60 of the unsuccessful problem solvers answered item 14 incorrectly, compared to 36 out of 60 of the successful problem solvers. The Teacher's Guide of the STEP showed this test item to be the most difficult of our ten common items. In contrast, for item 10 (the easiest item) 21 unsuccessful problem solvers and only 4 successful problem solvers missed the item. A Spearman rank difference correlation between (a) a ranking of the difficulty levels and (b) a ranking of the discrepancies between the number of successful and unsuccessful students missing each item was computed and found to be .69.

Table 7

SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE B SUBJECTS ON FIVE ITEMS FOR GRADES 7, 8, 9

	SUCCESSFUL		UNSUCCESSFUL	
	Median	Range	Median	Range
Category 1 Possession of information				
Grade 7	21.5	19-25	15	6-21
Grade 8	22	19-24	19.5	15-24
Grade 9	24.5	21-25	21.5	17-24
Significance level* --	$p. < .02$	$p. < .01$		
Category 2 Retrieval of information				
Grade 7	21.5	18-25	12.5	6-21
Grade 8	21.5	16-24	18.5	13-25
Grade 9	23	16-25	20	13-23
Significance level --	$p. > .05$ (n.s.)		$p, < .02$	
Category 3 Extracting information				
Grade 7	21	19-25	13	8-20
Grade 8	22	18-24	18.5	15-23
Grade 9	22	19-25	19.5	14-23
Significance level --	$p. > .05$ (n.s.)		$p. < .01$	
Category 4 Combining operations				
Grade 7	20	12-24	13	6-20
Grade 8	21	15-24	16	12-22
Grade 9	21.5	15-25	15.5	12-23
Significance level --	$p. > .05$ (n.s.)		$p. > .05$ (n.s.)	
Category 5 Number of correct responses				
Grade 7	3	1-5	1.5	0-2
Grade 8	3	2-3	2	1-4
Grade 9	3	1-4	2	1-4
Significance level --	$p. > .05$ (n.s.)		$p. > .05$ (n.s.)	

*Kruskal-Wallis One-Way Analysis of Variance was used to test significance of differences among grade samples.

Table 8

SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL
SAMPLE B SUBJECTS ON FIVE ITEMS FOR GRADES 4, 5, 6

	SUCCESSFUL		UNSUCCESSFUL	
	Median	Range	Median	Range
Category 1 Possession of information				
Grade 4	22	19-25	21.5	19-25
Grade 5	25	20-25	20.5	5-23
Grade 6	25	20-25	23.5	17-25
Significance level* --	$p. < .02$		$p. < .05$	
Category 2 Retrieval of information				
Grade 4	19.5	17-24	20.5	17-24
Grade 5	24	20-25	18	5-24
Grade 6	23.5	17-25	20.5	15-25
Significance level --	$p. < .05$		$p. > .05$ (n.s.)	
Category 3 Extracting information				
Grade 4	22	19-25	21	18-25
Grade 5	25	20-25	20.5	5-25
Grade 6	25	20-25	23.5	15-25
Significance level --	$p. > .05$ (n.s.)		$p. > .05$ (n.s.)	
Category 4 Combining operations				
Grade 4	20	15-25	18.5	14-23
Grade 5	24	17-25	16.5	5-24
Grade 6	24	17-25	20.5	14-25
Significance level --	$p. > .05$ (n.s.)		$p. > .05$ (n.s.)	
Category 5 Number of correct responses				
Grade 4	3.5	2-5	3	2-4
Grade 5	5	2-5	3	1-4
Grade 6	4	3-5	3	0-5
Significance level --	$p. > .05$ (n.s.)		$p. > .05$ (n.s.)	

***Kruskal-Wallis One-Way Analysis of Variance was used to test significance of differences among grade samples.**

Table 9

Coefficients (rho) of Inter-Rater Correlation
on all Items, all categories, Sample C

Grade	N	Category 1 Possession	 2 Retrieval	 3 Extraction	 4 Combining	 5 Total
4	20	.716	.665	.741	.924	.722
5	20	.862	.583	.815	.844	.814
6	20	.839	.669	.753	.937	.845
7	20	.860	.885	.935	.898	.905
8	20	.843	.838	.894	.961	.909
9	20	.847	.745	.891	.882	.866

Differences between successful and unsuccessful

Table 10 contains a summary of scores obtained by successful and unsuccessful Sample C subjects on the ten common science test items. Considerable overlap was found in the score distributions between successful and unsuccessful subjects in all grade-categories.

The null hypothesis for Hypothesis 1 is: There is no difference between children classified as successful problem solvers and children, of comparable grade level, classified as unsuccessful problem solvers in measures of (a) possession of information, (b) retrieval of information, (c) extracting information, and (d) combining operations, when science achievement test items are used in interviews as a basis for obtaining measures. The Mann-Whitney U test was used to test this hypothesis. The level of confidence was set at .05. A one-tailed test was used since there was a basis for predicting direction. Table 10 reports the significance for the grade-category comparisons as well as for the number of correct answers given. All differences were found to be significant.

It will be noted in Table 10 that the differences between the performances of the successful and unsuccessful subjects found in the "extracting of information" and the "combining operations" categories, are relatively greater for Sample C than for Sample B. This may be due to differences in the characteristics of Samples B and C test items. Success in answering the social studies test items was more often contingent upon an understanding of terms such as "natural environment," "latitude," "lowest elevation above sea level," and so forth than was the case with the science test items. All the information required to answer a test item was included in the statements of some of the science test items. Success appeared to be contingent upon a subject's extracting and structuring information given.

A more systematic effort was made in the Sample C interviews to find out whether or not a subject understood the question asked in each test item. Table 11 details the number of test items rated as "not understood" within successful and unsuccessful grade samples. The table indicates that unsuccessful subjects were unable to understand some of the questions posed by the test items, while the successful subjects appeared to experience less difficulty in understanding what question was asked. The contrast is

Table 10

SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE C SUBJECTS ON TEN COMMON ITEMS

	SUCCESSFUL		UNSUCCESSFUL		Significance level of differences*
	Median	Range	Median	Range	
Category 1 Possession of information					
Grade 4	33	27-39	30	25-35	.05
Grade 5	34	29-38	28	22-40	.05
Grade 6	34.5	31-39	31.5	23-39	.025
Grade 7	37	34-42	31	26-39	.001
Grade 8	40	37-42	32.5	30-40	.001
Grade 9	40.5	34-42	35	26-39	.001
Category 2 Retrieval of information					
Grade 4	31.5	27-43	28.5	22-33	.01
Grade 5	34	25-37	24.5	20-41	.025
Grade 6	34	28-40	28.5	23-38	.01
Grade 7	37.5	35-41	29	24-38	.001
Grade 8	39.5	34-41	31	26-39	.001
Grade 9	39	34-42	35	25-40	.01
Category 3 Extracting information					
Grade 4	39	25-49	31	20-43	.01
Grade 5	41	31-47	27	15-47	.01
Grade 6	45.5	39-48	38	27-45	.01
Grade 7	47.5	37-50	35.5	31-47	.001
Grade 8	48	43-50	38.5	27-43	.001
Grade 9	49	39-50	42.5	22-48	.001
Category 4 Combining operations					
Grade 4	33	22-43	21.5	14-33	.01
Grade 5	35	28-41	20	15-42	.01
Grade 6	42	30-45	26.5	22-41	.001
Grade 7	45.5	33-50	30	27-40	.001
Grade 8	45.5	38-50	31	25-41	.001
Grade 9	47	33-50	36	19-45	.001
Category 5 No. of correct answers					
Grade 4	7	4-10	4	2-6	.01
Grade 5	6	3-9	5	2-8	.025
Grade 6	8.5	4-9	5	3-7	.01
Grade 7	9	6-10	6.5	4-9	.01
Grade 8	9	5-10	6	4-9	.001
Grade 9	9.5	4-10	6.5	6-9	.01

*Mann-Whitney U tests based on all items

Table 11

AVERAGE NUMBER OF TEST ITEMS "NOT UNDERSTOOD" BY SUCCESSFUL AND UNSUCCESSFUL SUBJECTS, (BY GRADES, SAMPLE C)

	Grades					
	4	5	6	7	8	9
Successful	2.7	1.6	1.1	.6	.8	.7
Unsuccessful	3.7	4.4	4.2	4.8	4.8	4.5

particularly striking at the grades seven, eight, and nine levels. For example, the average number of test items not understood by the ten successful grade seven subjects was .6; the average for the unsuccessful was 4.8. Thus, on the average, the question posed in nearly one-third of the fifteen items was not understood by the unsuccessful grade seven subjects. Among the grades four, five and six unsuccessful problem solvers, no evidence of progress was found in the ability to understand the question posed by test items. Some improvement may be noted for the successful groups.

The series of questions based upon the temperature chart and thermometer diagrams presented particular difficulty to the unsuccessful problem solvers. Question 22 (See Appendix II, p. 12), for example, required the student to extract information from the thermometer diagrams and locate a best fit for the extracted information in the temperature chart. Inability to understand what was required in the test item seemed to interfere with the ability of the student to extract and interpret the relevant information and structure the pieces of information to the demands of the question. The following extracts are illustrative.

Successful 5th grade subject

S: (reads question 22) Do you want me to answer both of the thermometers? (pause) Would that be for both of the questions? (pause) 80 degrees in the outdoor thermometer would be 12 noon and for the indoor thermometer it would be 72...uh (pause)... 12 noon.
Int: What letter?
S: G.

(Later in the probing portion)

Int: Tell me what number 22 is asking.
S: ...to find out what time of day the thermometers show...
Int: Why did you pick G? How did you arrive at your answer?
S: I looked at the thermometer.
Int: Then what did you do?
S: I looked at the temperature chart.
Int: Then what did you do?
S: I decided that the answer would be 12 noon...80 degrees outside and 72 degrees inside.

Unsuccessful 5th grade subject

Int: What does Question 22 ask you to do?
S: (reads the question...pause)

Int: Do you know?
S: (pause)
Int: (Interviewer rephrases the question--child still does not answer) Tell me how you got your answer.
S: (pause) I don't know.
Int: What is the reading on the indoor thermometer?
S: (pause)...(interviewer repeats the question)..uh..70.
Int: What is the reading on the outdoor thermometer?
S: (pause)...(interviewer repeats the question)...80.
Int: Do you know how you could tell me the time of day now?
S: No.
Int: Is there anything on the page that would tell you when the thermometers looked like that?
S: No...

Changing the mode of representation of a test item did affect its difficulty level. Question 1, Part I of Form 4B of the Science STEP test presents four pictures showing changed positions of the shadows of flagpoles. The question asks which of the four pictures shows the shadow at noontime. Subjects in grades four, five and six were given a golf tee and a small flashlight to aid them in visualizing shadows cast by the flagpole. Twenty-eight of the forty subjects who were unable to cope with the problem in its symbolic form were able to demonstrate an understanding when presented with the visual aids.

A number of items included in the interview tested the student's understanding of the nature of a scientific experiment. An increase in the number of students demonstrating an understanding of the scientific method occurred from grades four through seven. For Questions 12 and 19 of Part I of Form 4B of the Science STEP test, 20 percent of the responses of fourth graders; 37 percent of the fifth graders, 47 percent of sixth graders, and 60 percent of seventh graders revealed some understanding of the scientific method. The contrast between successful and unsuccessful subjects in grades four, five and six was not great. On the other hand, a marked improvement by grade in the understanding of the scientific method did not occur between grades seven and nine. Responses to questions designed to probe the student's understanding of the scientific method did reveal a marked superiority of successful over unsuccessful grades seven, eight, and nine problem solvers. For successful problem solvers, 82 percent, while for the unsuccessful problem solvers 43 percent of the responses revealed some understanding of the scientific method.

As for Sample B, a comparison was made of the performances of successful and unsuccessful problem solvers on selected test items which both initially answered incorrectly. The basis for comparison on Sample C is not as large as on Sample B. While items were selected which were missed most frequently by successful subjects, the rate of misses by successful subjects was much lower on Sample C than on Sample B. A total of 119 incorrect responses were recorded for successful subjects on Sample C, while 224 incorrect responses were recorded for successful subjects on Sample B. The number of incorrect responses by unsuccessful subjects was also less on Sample C, with 188 incorrect responses to 270 incorrect responses on Sample B. On Sample C, the successful problem solver again performed somewhat better on each of the four categories in almost every instance. Table 5 contains the average ratings of successful and unsuccessful problem solvers on a 5-point rating scale.

Ratings are shown for all grades and an all-grade total is also presented. It will be noted that the average rating is low, even for the successful pupils. However, the successful pupils performed somewhat better than did the unsuccessful pupils, with differences in ratings ranging from .63 on Category 3 to .14 on Category 2. It will be noted that the greatest range on Sample C occurs in the category showing the smallest range on Sample B.

On the basis of an examination of protocols of both successful and unsuccessful problem solvers, some generalizations can be made.

As on Sample B, successful problem solvers who initially answered the question incorrectly tended to change to the correct answer with little or no probing or help from the interviewer. The same was not true to the same extent for the unsuccessful problem solver, who reached the correct answer, if at all, only after considerable probing and discussion with the interviewer. The following excerpts serve to illustrate this point:

Successful 9th grade subject
Int: O.K., now 22, what does the question ask you to do?
S: To tell the time that the temperatures are shown.
Int: How did you get your answer?
S: Well, I looked at the temperature on the indoors and it looked about 71 and I looked on the chart to see which questions has a temperature of 71 and I looked at the outdoor then and saw it was 80 and saw...oh..
Int: You said 10. Do you want to change?
S: Yes, to 12 noon, because uh...the outdoor temperature was 80 and the indoor was 72.

Unsuccessful 9th grade subject
Int: What does 22 ask you to do?
S: uh...(reads question)...to me it asks...It tells the degrees and it asks which temperature is drawn here.
Int: How did you get F for your answer?
S: Well, F, indoor was 80 and outdoors was 71, but the indoor thermometer doesn't go up to 80 degrees and the others weren't on the thermometer.

The preceding excerpts also illustrate the tendency of the successful problem solver to structure a problem more efficiently than the unsuccessful problem solver, even when both miss the question--a point which was also noted in Sample B.

The successful problem solver was better able to take advantage of rephrasing and focusing on a question than was the unsuccessful problem solver.

Successful 9th grade subject
Int: What is the reading of the indoor thermometer?
S: 72.
Int: The outdoor thermometer?
S: 80.
Int: If this were the indoor and this the outdoor what time would it be?
S: About 10.

Unsuccessful 9th grade subject

Int: O.K., about what temperature is the indoor thermometer?

S: 71.

Int: The outdoor?

S: 80.

Int: If this were the indoor and this the outdoor, what time would it be?

S: (pause)...well, there wouldn't be an answer then cause then it would go to 80 on the indoor and it wouldn't be a question...I couldn't think of one...

Int: If you had to pick one, which one would you pick?

S: I don't think I'd choose any of them.

The successful problem solver tends to be able to recognize common elements in a problem to a greater extent than does the unsuccessful problem solver.

Successful 6th grade subject (See Appendix II, p. 23)

Int: O.K., let's go to number 7.

S: (Reads question 7)...uh...I think it would be B (reads B).. because since the others didn't dry and the first and second one did and the first one could get sunlight and the second one was put where just air could get to it and not sunlight..no..I think it would be A...

Int: What do 1 and 2 have in common?

S: Both of them could get air.

Int: What do 3 and 4 have in common?

S: Neither one of them could get air.

Unsuccessful 7th grade subject

Int: O.K., let's go to Question 7.

S: (Reads question)...it would be B.

Int: ...Why B?

S: uh...cause 2...on 2 he just placed it where it could get uh..uh ...number 2 didn't get anything but air to dry it.

Int: What do 1 and 2 have in common?

S: They're both painted.

Int: What do 3 and 4 have in common?

S: The same thing as 1 and 2, I guess.

There appears to be some tendency on the part of the successful problem solver to be more able to deal with a problem from another frame of reference than is true for the unsuccessful problem solver.

Successful 9th grade subject (See Appendix II, p. 25)

Int: O.K., now 20.

S: (Reads question 20)...uh...I'd say E..

Int: ...O.K., would you please answer the question again as if you were a scientist?

S: uh....if I was a scientific person I would use F.

Unsuccessful 9th grade subject

Int: O.K., now 20.

S: (Reads Question 20)...G

Int: ...Would you answer the question again as if you were a scientist?

S: E.
Int: O.K., why not F?
S: 'Cause you just can't go to the moon and plant seeds and nobody has been to the moon.

Differences between grades

An inspection of Table 10 reveals that children in the upper grade levels appear to do somewhat better than children in the lower grade levels on each major category and on the number of items answered correctly. This was also the case in Samples A and B. The Kruskal-Wallis One-Way Analysis of Variance was used to test whether the scores from the different grades could have been drawn from the same population (Hypotheses 2 and 3). The null hypothesis was rejected at the .05 level or better in all categories for the successful problem solvers. The null hypothesis was accepted in categories 1 and 2 and rejected in categories 3, 4, and 5 for unsuccessful problem solvers. Table 12 summarizes these findings.

Table 12

SAMPLE C KRUSKAL-WALLIS ONE-WAY ANALYSIS OF VARIANCE
BETWEEN GRADES (TEN COMMON ITEMS)

		Level of Significance
Category 1	Possession of information	
Successful		.001
Unsuccessful		*
Category 2	Retrieval of information	
Successful		.01
Unsuccessful		*
Category 3	Extracting information	
Successful		.001
Unsuccessful		.05
Category 4	Combining operations	
Successful		.001
Unsuccessful		.001
Category 5	Number of correct answers	
Successful		.01
Unsuccessful		.01

*not significant at the .05 level.

It may be noted, from Table 10, that the medians do not increase by grade in a uniform manner. For the successful problem solvers, gains occur up to grade seven. The grades eight and nine pupils, in general, did little better than the grade seven pupils. The maximum median in each category is 50. The attained medians in categories 1 and 2 were not so high that continued increasing improvement could not have occurred. However, very little improvement occurred apparently for grades eight and nine in categories 1 and 2. On the other hand, in categories 3, 4, and 5 the attained medians approach the maximum possible median of 50. Therefore, as was the case in Sample B, the lack of improvement may be attributable to the level of the difficulty of the test items included in the interviews.

The significance of differences among grades in performances on each category and for total number of correct responses for the five items given to grades four, five and six and to the five items given to grades seven, eight and nine successful and unsuccessful subjects was tested by the Kruskal-Wallis One-Way Analysis of Variance. None of the differences among grades was found to be significant at or beyond the .05 level. This may be attributable to a number of possible factors. Five items may be too small a number of items to detect the progress that actually does occur within a three grade range. The items may be too easy for the successful problem solver to leave room for growth at the upper grade levels. Unsuccessful problem solvers may make little progress as they proceed from grade to grade. To better assess the factors contributing to the apparently inadequate progress from grade to grade, Mann-Whitney U tests were applied to test the significance of differences between the successful and the unsuccessful subjects for each grade category. Table 13 presents a summary of the data used in applying the Mann-Whitney U tests and the significance level of the differences found among grades four, five and six subjects. Table 14 presents similar information for grades seven, eight and nine subjects.

Differences between grade-category scores of groups of successful and unsuccessful problem solvers were not significant for grades four and five for the five items given only to grades four, five and six children (Table 13). From Table 10, the comparable grade-category distributions of scores for grades four and five on the ten common.items were significantly different. But, the same five items did function to discriminate between the successful and unsuccessful at the sixth grade level. All but two of the differences between grade category scores of successful and unsuccessful problem solvers were significant for grades seven, eight and nine on the five non-common items. This would suggest that these items function better in discriminating between the successful and the unsuccessful than they do in discriminating between grade-level performances.

Hypotheses 2 and 3, concerning superior performances of higher grade level children over lower grade level children, in general, are supported by the data, although the differences among grades is not as clear-cut as the differences between successful and unsuccessful groups of students.

Table 13

SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE C SUBJECTS FOR GRADES 4, 5, AND 6 ON FIVE NON-COMMON ITEMS

	SUCCESSFUL		UNSUCCESSFUL		
	Median	Range	Median	Range	Significance level of differences*
Category 1					
Possession of information					
Grade 4	17.0	11-23	16.5	11-19	n.s
Grade 5	17.5	14-22	14.5	9-22	n.s.
Grade 6	21.0	13-23	15.0	11-22	.01
Category 2					
Retrieval of information					
Grade 4	16.0	8-22	14.5	9-18	n.s.
Grade 5	17.0	12-22	13.5	8-20	n.s.
Grade 6	21.0	14-23	12.5	11-19	.01
Category 3					
Extracting information					
Grade 4	21.5	19-25	20.5	18-23	n.s.
Grade 5	22.5	18-25	21.0	16-25	n.s.
Grade 6	23.5	20-25	20.0	17-24	.025
Category 4					
Combining operations					
Grade 4	18.0	8-25	14.0	8-20	n.s.
Grade 5	20.0	14-25	15.0	9-23	.05
Grade 6	23.0	16-25	14.5	9-18	.001
Category 5					
No. of correct answers					
Grade 4	3.0	1-5	3.0	1-5	n.s.
Grade 5	3.0	0-5	2.0	1-5	n.s.
Grade 6	4.0	3-5	2.5	1-3	.001

*Mann-Whitney U test used to test significance of differences.

Table 14

SUMMARY OF SCORES OBTAINED BY SUCCESSFUL AND UNSUCCESSFUL SAMPLE C SUBJECTS FOR GRADES 7, 8, AND 9 ON FIVE NON-COMMON ITEMS

	SUCCESSFUL		UNSUCCESSFUL		
	Median	Range	Median	Range	Significance level of differences*
Category 1 Possession of information					
Grade 7	20	16-23	12.5	11-17	.001
Grade 8	21	17-22	13.5	9-17	.001
Grade 9	21	20-23	16	9-19	.001
Category 2 Retrieval of information					
Grade 7	17	16-17	14.5	13-17	.01
Grade 8	17	15-17	16	11-17	.05
Grade 8	17	16-17	17	13-17	n.s.
Category 3 Extracting information					
Grade 7	23.5	20-25	17.5	13-23	.01
Grade 8	23.5	19-24	18	11-24	.01
Grade 9	23	21-24	21	11-23	.001
Category 4 Combining operations					
Grade 7	21.5	15-25	13.5	9-24	.01
Grade 8	23	18-25	12.5	7-20	.001
Grade 9	21	18-25	17.5	9-23	.05
Category 5 No. of correct answers					
Grade 7	4	2-5	2.3	0-4	.05
Grade 8	4	3-5	4	2-5	.001
Grade 9	4	2-5	2.3	1-5	n.s.

*Mann-Whitney U test used to test significance of differences.

Patterns of Weaknesses of Unsuccessful Problem Solvers

In order to determine whether or not the difficulties which unsuccessful problem solvers experienced, followed any particular pattern of weaknesses, a chart was made for each grade showing, for each unsuccessful and each successful subject: (1) his total score for each of the four categories, (2) the corresponding rank assigned to this score when compared with ranks received by all other subjects in his grade, (3) a total numerical score of the scores received in the four categories, (4) the total number of questions answered correctly, initially, (5) the rank of the number of correct responses when compared to all other students in his grade and (6) the rank of the total score in all categories when compared to all other subjects in his grade. Charts of this kind were made for Samples A, B, and C, grades four through nine.

The rank at which each student fell in each of the categories was compared with his ranks in each of the other categories, in an effort to ascertain the area or areas where he was experiencing difficulty or weakness in the problem-solving process. A student was considered to be relatively strong or weak in a category whenever his performance in that category ranked four or more ranks higher or lower than his ranked performance in one or more other categories. For instance, observing the rankings received by the first and second students represented below, we may assume that although the quality of the second student's performance is certainly superior to that of the first, neither student can be said to have experienced a great deal more difficulty in any one category or area of the problem-solving process than in any other category or area.

	Rank of Student 1	Rank of Student 2	Rank of Student 3
Category 1	2.0	11.5	6.5
Category 2	2.0	11.0	6.0
Category 3	1.5	11.5	1.0
Category 4	1.5	11.0	1.0
Total 1-4	1.0	9.0	1.0
No. of correct answers	3	11.0	2.5

On the other hand, the third student placed more than four rankings higher in categories one and two than he did in categories three and four. It may, therefore, be assumed that this student was relatively stronger in (1) possession of information and (2) retrieval, than in (3) extracting and (4) combining operations.* This student could be said to have a pattern of weaknesses in categories one and two, with relatively less weakness in categories three and four. From similar analyses of individual patterns of all students in Samples B and C, fifteen patterns of weakness were identified and the number of students falling into each pattern was recorded (See Tables 15 and 16). It is interesting to note, that of the fifteen possible combinations of the four categories, at least one or more of the unsuccessful problem solvers in Samples B and C possessed one of the fifteen combinations or patterns of weaknesses. This would tend to agree with Buswell and Kersh (1956) who came to the conclusion "that either (a) great variety in the process of problem solving is normal and to be expected or (b) schools have been ineffective in teaching a technique of problem solving..." (1956, pp. 136-137).

---

*Poorest performance in a grade was ranked 1; the best performance was ranked 20.

It was observed that, in each grade and in each category, a few unsuccessful problem solvers ranked higher than some successful problem solvers. Also, in each grade and in each category, a few unsuccessful problem solvers ranked at or above the median for that category in that grade. In order to determine the extent to which this occurred, a chart was made for Samples B and C showing the number of unsuccessful problem solvers in each grade whose rank in each category fell at or below the median rank for that grade. In no category, and in no grade, were there less than seventy percent of the unsuccessful problem solvers who fell below the median in any category. From seventy to one hundred percent of all unsuccessful problem solvers in each grade were at or below the median for that grade (1) in each of the four categories scores, (2) in the rank of the total number of correct responses, and, (3) in the rank of the total of the scores for all four categories.

While there were very few unsuccessful problem solvers who were consistently above the median for their grade in all categories--an average of less than one student per grade in both Samples B and C--records like the one for the first student below (#174) did exist. Much more frequently, an unsuccessful problem solver who ranked at or above the median for his grade in one or more categories, ranked below the median in other categories (e.g., Student #397 below). By far the most frequent pattern which occurred among the unsuccessful problem solvers was a very low rank on all categories and areas, as shown by the record of the third student below (#310).

		Ranks	
	Student 1 (#174)	Student 2 (#397)	Student 3 (#310)
Category 1	13.5	15.0	1.0
Category 2	12.5	7.5	2.0
Category 3	14.0	10.0	1.5
Category 4	12.0	12.5	2.0
Total 1-4	12.5	11.0	2.0
No. of correct responses	13.0	6.0	2.5

In every grade, in both Sample B and Sample C, there appeared to be several unsuccessful problem solvers who were apparently completely "lost" in their problem-solving skills--the two or three students in each grade who received the lowest total ranks, either by fewest number of correct responses and/or by the lowest total score on the four categories, were in the majority of cases, consistently the lowest in all of the categories ranked, as was student #310 shown above. As could be expected, this situation also occurred with those two or three successful problem solvers who consistently ranked the highest in all areas and categories. It was those students, both successful and unsuccessful, whose total ranks were near the median rank for their grade level, who showed the greatest variation in their ranks between the various categories--which probably indicates that it was these students whose patterns of strengths and/or weaknesses also

Table 15

PATTERNS OF WEAKNESSES OF SAMPLE B UNSUCCESSFUL SUBJECTS

	Strength in Category		Number in Grade							
	Strong	Weak	4	5	6	7	8	9	Total	Percent
1.	none	1, 2, 3, 4*	3	5	3	6	5	3	25	41.66
2.	1, 2*	3, 4	1	3	0	2	0	1	7	11.66
3.	1	2, 3, 4	0	0	0	0	0	0	0	0
4.	1, 2, 3	4	1	0	0	0	0	0	1	1.66
5.	1, 3	2, 4	1	0	1	1	0	0	3	5.0
6.	3	1, 2, 4	0	0	1	0	1	2	4	6.66
7.	1, 4	2, 3	0	0	0	0	1	0	1	1.66
8.	2, 4	1, 3	1	0	0	1	0	1	3	5.0
9.	1, 3, 4	2	0	1	1	0	0	1	3	5.0
10.	1, 2, 4	3	0	0	1	0	1	1	3	5.0
11.	2	1, 3, 4	0	1	1	0	0	0	2	3.33
12.	2, 3	1, 4	2	0	1	0	0	0	3	5.0
13.	2, 3, 4	1	0	0	0	0	1	1	2	3.33
14.	4	1, 2, 3	0	0	1	0	0	0	1	1.66
15.	3, 4	1, 2	1	0	0	0	1	0	2	3.33
Total			10	10	10	10	10	10	60	

Table 16

PATTERNS OF WEAKNESSES OF SAMPLE C UNSUCCESSFUL SUBJECTS

	Strength in Category		Number in Grade							
	Strong	Weak	4	5	6	7	8	9	Total	Percent
1.	none	1, 2, 3, 4*	3	3	5	4	4	5	24	40.0
2.	1, 2*	3, 4	1	1	3	2	0	0	7	11.66
3.	1	2, 3, 4	1	0	1	0	1	0	3	5.0
4.	1, 2, 3	4	2	0	0	0	1	0	3	5.0
5.	1, 3	2, 4	0	0	0	0	0	1	1	1.66
6.	3	1, 2, 4	1	1	1	0	0	0	3	5.0
7.	1, 4	2, 3	1	0	0	1	0	0	2	3.33
8.	2, 4	1, 3	0	0	0	0	1	0	1	1.66
9.	1 3, 4	2	0	0	0	0	1	0	1	1.66
10.	1 2, 4	3	0	1	0	0	0	0	1	1.66
11.	2	1, 3, 4	0	0	0	0	2	1	3	5.0
12.	2, 3	1, 4	0	1	0	0	0	1	2	3.33
13.	2, 3, 4	1	1	1	0	1	0	0	3	5.0
14.	4	1, 2, 3	0	0	0	0	0	1	1	1.66
15.	3, 4	1, 2	0	2	0	2	0	1	5	8.33
Total			10	10	10	10	10	10	60	

*1. Possession of information
2. Retrieval of information
3. Extracting information
4. Combining operations

demonstrated marked differences. An observation of the records of the individual students in Samples B and C whose total scores fell near the median for their grade showed that their patterns of strengths and weaknesses did vary substantially. For example, among the forty-nine unsuccessful problem solvers in Samples B and C who exhibited patterns of general overall weaknesses in all categories, only eleven of these students were students whose total scores ranked near the median rank for their grade. Thirty-three of these forty-nine unsuccessful problem solvers with patterns of overall general weaknesses ranked extremely low (in the bottom twenty percent of successful and unsuccessful students) in either total number of correct responses and/or total score on all four categories. The remaining five unsuccessful problem solvers with overall general weakness patterns ranked well above the median for their grade in one of the total rankings, either total category score or number of correct responses.

Variability in performance on each of the four categories characterized the student who was closer to average in overall performance. It might be anticipated, therefore, that more variability in category scores, than was found in this study, would occur in a study that included average problem solvers as well as successful and unsuccessful problem solvers.

The overall weakness pattern

A reexamination of the protocols of students weak in all categories revealed that the degree of weakness in the problem-solving processes varied considerably among these students. The total number of questions answered correctly did not discriminate between the students who were "hopelessly lost" and the students who were able to perform fairly well in some areas of the problem-solving process, and yet were unable to reach a thoroughly satisfactory answer to the problem. For example, the following material, taken from protocols of two unsuccessful problem solvers who failed initially to answer the same question correctly during the Sample C interviews, illustrates the qualitative difference between the answers of the two subjects. While both students experienced difficulty with the problem, the first student was not able to profit from the probing session, whereas, the second student was eventually able to clarify his thinking and reach the correct solution.

Each subject read aloud and answered the following question:

"The sun is the only object in our solar system that makes and gives off large amounts of heat and light. This means that we see the moon because it
A reflects light from the sun
B gives off its own light
C is not in our solar system
D reflects much light from other planets."

First subject

S: (reads question)--I think D is right.
Int: What does the question ask you to do?
S: (pause)--it asks me..pause...
Int: What?
S: (long pause)--I don't understand--
Int: Look at it again, and tell me what the question asks you to do.
S: It's asking me about some objects in the solar system---

2

Int: Why did you select D as your answer?
S: I don't know---it just sounds better.
Int: Why?
S: (pause) It just does.
Int: What's a solar system?
S: --(pause)--(interviewer repeats question) Something on the earth.
Int: Tell me what you mean.
S: I don't know.
Int: What is our solar system?
S: The earth.
Int: Anything else?
S: The middle of the earth.
Int: Is the moon in our solar system?
S: No--I don't think so--I can't remember.
Int: What is the differences between a planet and a star?
S: A planet is bigger than a star.
Int: Any other difference?
S: A planet is big, a star has points.
Int: Is the sun a planet or a star?
S: A planet.
Int: Do planets give off large amounts of heat and light?
S: No.
Int: Does the earth give off large amounts of heat and light?
S: Yes.
Int: Why?
S: We have electricity.
Int: Does the moon give off its own light or reflected light?
S: Gives off its own light.
Int: Where does the light in our solar system come from?
S: The moon and the sun.
Int: Why did you not choose A as your answer?
S: Light on the moon is not reflected from the sun.
Int: Why not C?
S: Moon is not in our solar system.
Int: Why not D?
S: Moon doesn't reflect light from other planets.

Second subject

S: (reads question)--B is the answer.
Int: What does the question ask you to do?
S: --uh--means that we see the moon because it gets light from the earth.
Int: Is that what it asks?
S: I'm trying to find out if it's answer A, B, C, or D.
Int: But what does the question ask you to do?
S: It asks me--(pause)--we see the--oh, I know--why do we see the moon?
Int: Why did you select the answer you did?
S: I'm not sure.
Int: What is a solar system?
S: uh--(pause)--uh--it's just open space.
Int: What is our solar system?
S: ---pause---I don't know.
Int: Is the moon in our solar system?
S: Yes.

Int: What is the difference between a planet and a star?
S: Well--a star is--sort of like the sun, you know--and a planet, well, it isn't like--it's kind of cooled off--like the earth.
Int: Is the sun a planet or a star?
S: Star.
Int: Do planets give off large amounts of heat and light?
S: No.
Int: Does the earth give off large amounts of heat and light?
S: No.
Int: Why not?
S: It's a planet, I guess.
Int: Does the moon give off its own light or reflected light?
S: Reflected.
Int: Where does the light in our solar system come from?
S: The sun.
Int: Why did you not select A as your answer?
S: Oh, A *is right*.
Int: Do you want to change your answer from B to A?
S: Yes.
Int: Why?
S: 'Cause the moon *does* get its light from the sun.
Int: Why not B?
S: 'Cause the moon *doesn't* give off its own light.
Int: Why not C?
S: 'Cause the moon *is* in our solar system.
Int: Why not D?
S: 'Cause it doesn't reflect light from the other planets.

In order to determine whether or not the generalizations which have been made concerning patterns of weaknesses which unsuccessful problem solvers demonstrate when solving problems in social studies will also hold true for unsuccessful problem solvers when they are solving problems in science, the records of unsuccessful problem solvers who were interviewed first with social studies materials and later with science materials were examined and compared, using the same criteria as was used previously for identifying patterns of weakness. Eighteen of the thirty unsuccessful problem solvers in Sample A (who were also interviewed as a part of Sample C) exhibited patterns of general overall weaknesses in all categories of the problem-solving process. Sixteen of these same thirty unsuccessful problem solvers from Sample A showed patterns of general overall weaknesses in all categories in their Sample C interviews. This suggests that at least this one major weakness pattern may carry over from one subject-matter area to another. The record of ranks for the performance of subject number 217 shown below, is representative of these students who demonstrated similar patterns in both interviews.

Rank for Student #217

	Sample A	Sample C
Category 1	1.0	2.0
Category 2	2.0	2.0
Category 3	2.0	1.5
Category 4	1.5	1.5
Total 1-4	6	7
No. of correct answers	4.5	5.5

Asking the "right" questions

The protocol material for some students seems to suggest that skill in problem solving involves a skill in "asking oneself the right questions" when confronted with a problem situation. During the interview sessions, the interviewer might have served as a model to some students with respect to "asking oneself the right questions." At any rate, towards the end of an interview some students themselves were raising the same type of questions about the problems as those included in the probing section of the interview process. On earlier questions, they were, as it were, "catching on" to the types of questions one needs to raise in coping with problem-solving type achievement test items.

For example, a fourth grade subject initially identified as an unsuccessful problem solver was interviewed in both Samples A and C. In the Sample A interview, he answered only three questions correctly and could not justify his answers to one of these problems. He seemed to lack stored information, had difficulty collapsing the question to its basic elements, could not extract relevant information and failed to demonstrate any useful problem-solving structuring.

His protocol materials for Sample C showed very little improvement in the number of correct responses, but as the Sample C interview progressed, the quality of his performance showed steady improvement. A part of the interview is included here. He read and answered the same question as is quoted above--"The sun is the only object in our solar system. . ."

S: It would be--uh--B is right answer.
Int: What does the question ask you to do?
S: uh--uh--to find out we see--oh, why we see the moon.
Int: O.K., why did you select the answer you did?
S: I selected B because-- no, B can't be right--'cause it says the sun only gives us light--well, D is wrong too, 'cause if only the sun gives off light it can't get light from other planets--it must be A--wait a minute, C says the moon is not in our solar system--uh--I guess it is in our solar system, if we can see it. The question tells you that--it's gotta be A, cause the moon gets it's light from the sun.
Int: Why not B?
S: It doesn't give off its own light--the sun shines on the moon--that's why we can see it, I guess.

This raises a question as to the kinds of things that the teacher needs to do to provide a good model in problem solving for the student.

Problem-solving process during interview

During the course of the interviews, the impression was gained that many students were improving in their problem-solving processes. As a rough check on this impression, scores obtained by those students who were interviewed on both the social studies materials and the science materials were compared. This group was composed of the thirty unsuccessful problem solvers (five from each grade, four through nine) who were a part of Sample A and who were also interviewed as a part of Sample C.

On a basis of chance alone, one would expect some deviation in performance, with some students doing better in social studies than in science and others performing relatively better in science. Most of the thirty unsuccessful problem solvers, however, who were interviewed during the social studies interviews (Sample A) improved in their problem-solving processes when interviewed for the second time with science problems (Sample C). Seventeen of the thirty unsuccessful problem solvers ranked relatively higher in science than in social studies on the number of questions answered correctly. Ten of the thirty unsuccessful problem solvers ranked relatively lower on the science test, and three students ranked the same on both tests. When ranked by the total score on all four categories, twenty-one of the thirty unsuccessful problem solvers ranked higher on the science set of problems, three ranked lower on the science set, and five students ranked the same on both tests. Thus, the number of students whose performance improved is large enough to suggest that some adaptation of the interview process may have promise in teaching problem-solving skills.

Limitations of Interview Technique

Some limitations in the technique utilized in this research study of the problem-solving processes of students became evident in the protocols of a few students. These students, in the non-probing portion of the interview, arrived at the correct answer to many or most of the test items; yet, in the probing session, were unable to provide satisfactory answers to the questions asked. This might not be due to an inordinate success in guessing. It may be due to an inability to verbalize or to communicate in providing explanations. For example, the record of one ninth grade unsuccessful problem solver given below, shows how he ranked on each of the four categories, and on total number of correct responses.

Rank	Student #407
Rank on Category 1	2.0
Rank on Category 2	2.0
Rank on Category 3	3.5
Rank on Category 4	3.0
Rank on Total 1-4	2.0
Rank by No. of correct answers	11.5

This student initially answered nine of the fifteen problems correctly, a number which placed him above the median for the number of correct responses in his grade. Upon examining the protocol materials for this student, it seemed apparent that this student experienced great difficulty in verbalizing explanations; he had trouble collapsing the problem into its basic elements; he seemed to understand what the question asked him to do, but could not explain it; he could not generalize from specific situations, or discuss the manner by which he arrived at the correct answer; he could not verbalize how he applied rules or principles to a particular problem; and his vocabulary was quite limited. A problem in communication no doubt contributed to this great discrepancy between the number of questions answered correctly and the apparent quality of his problem-solving process.

In Chapter V, a summary of the results obtained from the analysis of data is presented.

Chapter V

SUMMARY AND CONCLUSIONS

Objectives

The major objectives of this study were:

1. to explore the usefulness of the individual interview technique as a means of identifying and understanding the problem-solving processes of children as they worked on achievement test items of a problem-solving type;
2. to find out, through an analysis of protocols of testing interviews, what categories, constructs, and dimensions are useful in describing the differences between the problem-solving processes of the successful and unsuccessful problem solver and the differences among the problem-solving processes of children of different grade levels;
3. to test whether the categories, constructs, and dimensions of problem-solving processes found to be discriminating within one sample of successful and unsuccessful problem solvers also is found to be discriminating for a second sample, using the same test materials and for a third sample, using test materials from a different subject-matter area.

Procedure

Success in problem solving was assessed on the basis of performance on Form 3A or 4A of the Social Studies, Sequential Test of Educational Progress. About 120 children in each grade level, four through nine, were administered the appropriate grade-level form of this test. Children whose scores fell in the top third for their grade were identified as "successful" problem solvers; and children scoring in the bottom third for their grade were identified as "unsuccessful" problem solvers. Three samples were drawn from this pool of subjects. Each sample contained ten successful and ten unsuccessful problem solvers per grade. Samples A and B subjects were interviewed using test items in social studies. Science test items were used in Sample C interviews.

Fifteen test items from Form 3B and 4B of the Social Studies, Sequential Test of Educational Progress were administered to each subject. Ten items were included within the interviews of the children in all grade levels; five items from Form 4B of the STEP test were included for children in grade four, five, and six; and five items from 3B of the STEP test were included for children in grades seven, eight, and nine.

An Interview Guide, consisting of two portions, was developed for use in the Sample A interviews. In a non-probing portion of the interview, each child was presented with a test item and was encouraged to "think aloud" as he attempted to answer the item. In a probing portion of the interview, the interviewer attempted to determine what cognitive processes were involved as the child arrived at his answer. Prior to the interviews, the interviewers participated in a series of staff meetings. Various conceptualizations of the problem-solving process found in the literature were discussed and each test item to be included in the interview was discussed in the light of the various conceptualizations. A series of questions to be included in the probing portion of the Interview Guide developed out of the staff discussions. These questions were revised on the basis of the experience gained in the Sample A interviews. All interviews were tape recorded and transcribed.

An analysis of the protocols of Sample A interviews seemed to indicate that differences between successful and unsuccessful problem solvers in problem-solving processes, and differences among grade levels could be subsumed under four categories: (1) information possessed; (2) information retrieved; (3) extraction of information; and (4) combining operations. The Interview Guide was revised so as to focus questions more sharply on each category for Sample B interviews, and was adapted to science materials for Sample C interviews.

A scoring procedure was developed for the purpose of evaluating the protocols in terms of scores on each of the four categories. These scores constitute the data used to test hypotheses generated from the second and third objectives of this study.

Analysis of Data and Results

One hypothesis emerging out of an analysis of Sample A protocols was:

> Children, in grades four through nine, classified as successful problem solvers are superior to children of comparable grade level, classified as unsuccessful problem solvers in measures of (a) possession of information, (b) retrieval of information, (c) extracting information, and (d) combining operations.

The interview responses by each Sample A subject to each of the fifteen test items were rated on a five point scale for each of the four major categories. Ratings were summed to yield a total score on each category for each subject. The Mann-Whitney U test was applied to the score distributions of the twenty-four grade-category samples to estimate the probability that successful and unsuccessful grade-category score distributions could have been drawn from the same population. All differences were found to be significant beyond the .05 level. Successful problem solvers were superior in performance to unsuccessful problem solvers on all four categories in each of the six grades.

There was some overlap in the score distributions of successful and unsuccessful grade-category samples. Some unsuccessful subjects did better than some successful subjects. But this held true for the number of correct answers provided to the fifteen test items within the interviews by successful and unsuccessful subjects. This may have been due to faulty original classification of children into successful or unsuccessful problem solvers, or possibly the interview problem-solving situation differed significantly from the paper-and-pencil problem-solving situation for some children.

The four categories of the problem-solving process that were hypothesized as differentiating the successful from the unsuccessful problem solver were generated from two sources. First, the four categories seemed to subsume differences observed in the protocols of successful and unsuccessful Sample A subjects. Second, the observations of differences were made after a fairly intensive survey of the literature on problem solving. Protocols were read to see whether various categories, constructs, and dimensions discussed in the literature were helpful in differentiating the successful from the unsuccessful Sample A problem solvers. To determine whether the four categories that differentiated successful from unsuccessful Sample A subjects also differentiated for a fresh sample, data derived from Sample B protocols were analyzed in the same way as for Sample A. To determine whether the four categories were useful in differentiating in another subject-matter area--science--data derived from Sample C protocols also were analyzed in the same way as for Sample A.

Mann-Whitney U tests applied to Sample B data indicated that in nineteen of the twenty-four grade-category comparisons, successful subjects performed significantly better than unsuccessful subjects. In the remaining five comparisons (all grade four scores and the "extracting information" scores for grade six) successful problem solvers were superior but not significantly superior to unsuccessful problem solvers. Mann-Whitney U tests applied to Sample C data indicated that, in all twenty-four grade-category comparisons, successful subjects performed significantly better than unsuccessful subjects.

The poorer discrimination found between successful and unsuccessful Sample B subjects in grade four may have been due to the relatively greater difficulty experienced by the interviewers in eliciting verbalizations from grade four children. The inter-rater reliability coefficients on the grade four Sample B category scores and total process scores were relatively low. All but one was below .70. In comparison, twenty-one of the thirty inter-rater-grade reliability coefficients for Sample B and twenty-two for Sample C were above .80.

A second hypothesis emerging out of an analysis of Sample A protocols was:

> Children classified as successful problem solvers in the upper grade levels are superior to children classified as superior in the lower grade levels in measures of (a) possession of information, (b) retrieval of information, (c) extracting information, and (d) combining operations.

A third hypothesis parallels the second hypothesis for unsuccessful problem solvers.

Ten of the fifteen items included in the interviews were common for all grades, four through nine. The Kruskal-Wallis One-Way Analysis of Variance was used to test whether the scores (on these ten items) from the different grades (with successful and unsuccessful scores treated separately) could have been drawn from the same population. It was found that for Sample B, the null hypothesis of no difference could be rejected at the .05 level or better for all categories, for both successful and unsuccessful problem solvers. For Sample C, all differences among grades for successful problem solvers were significant at the .05 level or better. Among grade levels for unsuccessful problem solvers, differences were found to be significant at the .05 level or better except for the "possession of information" and the "retrieval of information" categories.

While Hypotheses 2 and 3 were supported by the data of this study, a further study of gains from grade to grade proved to be of interest. Medians did not increase by grade in a uniform manner. In both Samples B and C gains are rapid for the successful problem solvers up to grade seven. Grades eight and nine subjects did very little better than grade seven subjects. A study of the functioning of individual test items seemed to indicate that the difficult items functioned well in discriminating between students in successive grade levels while the easier items functioned well in discriminating between successful and unsuccessful problem solvers within a single grade level.

Kruskal-Wallis One-Way Analysis of Variance was used in analyzing differences among grades five and six and among score distributions of grades seven, eight, and nine on the four categories and in the total number of correct responses provided to the five non-common items. For Sample C, none of the differences among grades was found to be significant beyond the .05 level. For Sample B, in only one category ("information possessed") were differences found to be significant beyond the .05 level among grade samples of successful subjects. For Sample B unsuccessful subjects, significant differences were found on the first three categories.

In general, the differences found among grades in performances on each category and on total number of correct responses are not as clear-cut as are the differences between successful and unsuccessful subjects within a single grade. This may be due to characteristics of the test items included or to little actual improvement from grade to grade.

Something is lost when one attempts to reduce the contents of about 350 protocols of one-hour interviews into four category scores per protocol. In order to recapture some of the content, two additional questions were raised by the investigators. First, the question was raised as to what differences in the problem-solving processes of successful and unsuccessful subjects occurred on test items that were answered incorrectly by both successful and unsuccessful subjects. A comparison made of the performance of successful and unsuccessful problem solvers on selected items provided some support to the following generalizations.

(1) Successful problem solvers performed at a lower level on each of the four categories than they did on items which they answered correctly, but at a higher level than the unsuccessful problem solvers who failed the items. This generalization held true for both Samples B and C in almost every grade-category comparison.
(2) Successful problem solvers who initially answered the question incorrectly tended to change to the correct answer with little or no probing or help from the interviewer. Unsuccessful problem solvers were slower in reevaluating their responses.
(3) Successful problem solvers exhibited more facility in structuring the elements of a problem. This generalization held true particularly in items containing relatively large numbers of elements that needed to be structured.

(4) Successful problem solvers were better able to reason through a problem to some answer even when they had faulty information

(5) The successful problem solver recognized more than did the unsuccessful problem solver when he needed additional information.

(6) The successful problem solver was more apt in abstracting crucial similarities and differences in suggesting alternate ways of conducting a scientific experiment.

(7) The successful problem solver seemed to be better able to change his "set" and look at a problem from a changed frame of reference.

A second question raised was whether among unsuccessful problem solvers patterns of weaknesses could be detected. Fifteen patterns of weaknesses were found from an analysis of test scores of unsuccessful problem solvers in Samples B and C. About 40 percent of the unsuccessful problem solvers were weak in each of the four categories. About 30 percent of the unsuccessful students were relatively strong on "possession of information" and relatively weak on one or more of the remaining three categories. In general, the data in this study suggests that "possession of information" is usually a necessary but not a sufficient condition to insure success in solving an achievement test item.

A reexamination of the protocols of the unsuccessful problem solvers in Samples B and C provided some support to the following generalizations:

1. A considerable variation was found to exist in the quality of the problem-solving processes of subjects with an overall weakness pattern. But the total number of questions answered correctly did not discriminate adequately between the student who was "hopelessly lost" and the student who, with some help, was able to think constructively on a problem.
2. Success in answering a test item appeared to be contingent, to a considerable extent, upon "asking oneself the right questions" when confronted with a test question. Toward the end of an interview session, some students appeared to be asking themselves the same kind of questions that were included in the probing portion of the interview, and to be improving in their performance.
3. The problem-solving processes of students are amenable to improvement through instruction. During the course of the interviews, some students appeared to be improving in their problem-solving processes. A comparison was made of the process scores of the thirty unsuccessful problem solvers who were included in both Samples A and C. When ranked on the total score on all four categories twenty-one of the thirty unsuccessful problem solvers ranked higher on their second interview, three ranked lower, and five ranked the same.

Conclusions and Discussion

The categories, constructs, and dimensions that serve to describe problem-solving processes appear to vary according to the kind of problem tasks investigated and the methods employed in studying problem solving.

Even for the limited range of types of problems included in this study, the relative importance of our four categories varied from one test item to another. The relatively greater importance found in the information possessed by a student in this study as compared to the Bloom and Broder (1950) study may be due to differences in the characteristics of achievement test items in the two studies. For the type of problem employed by Rimoldi (1962, 1964) in his extensive research on problem solving, the ability to recognize what questions need to be answered in order to solve a problem constitutes a crucial ability.

The process of hypothesis formation and hypothesis testing did not emerge clearly in this study. In a restricted sense it may have been considered to occur as students evaluated alternatives in a multiple-choice question. On the other hand, hypothesis formation and testing, considered as a process of generating and developing a plan for evaluating alternative explanations for phenomana was not evidenced in the protocols. This may be attributed to the inadequacy of most achievement test items as a means of eliciting behavior as complex as hypothesis formation and testing.

The protocol materials in this study might have been analyzed in terms of TOTE units. The first "test" corresponds to "understanding of the question posed"; the "operate" category would include both "extracting information" and "combining operations" category; the second "test" constitutes one combining operation in which the answer reached is compared to the requirements of the problem; and the "exit" represents the student's request to move on to the next question.

There can be little doubt that non-cognitive factors affect the problem-solving processes. There were scattered indications in the protocols of this study that level of motivation, persistence, and attitudes are involved in the diverse ways children attack problems. However, in many of the protocols, such affective factors were difficult to assess.

The construct "structuring," appears to be useful in explaining the problem-solving processes reflected in the protocols of this study. Structuring seemed to occur in protocol materials relevant to all three process categories of this study. The structuring appeared to be of two kinds--first, the relationships that existed within the statement of the problem are perceived and comprehended; and second, the information contained in the question is related to the information the student possessed and retrieved.

The construct "search model" would seem to be particularly helpful in explaining processes underlying the "extraction of information" category. An analysis of the Sample C protocols revealed that unsuccessful problem solvers were particularly weak in understanding the question asked in a test item. It would seem obvious that searching behavior is facilitated when the problem solver understands what he is searching for.

One objective for this study was to explore the usefulness of the interview technique as a means of studying the problem-solving processes of children. While it seems possible to learn much in an interview about the process through which children solve problems, there are certain limitations to the technique. Thought is inferred from verbal responses. Poor ability of the problem solver to verbalize, thus, is one limitation of the interview technique.

A second limitation rests in the kind of inferences that can be made from the verbal responses to interview questions. It is not difficult to discover what information a student has extracted from the statement of a test question. It is more difficult to infer the nature of the "search model" that has directed his search for information. The essential nature of the kinds of associations or "train of thought" leading to a verbal response is difficult to infer with confidence from interview protocols.

However, despite the foregoing limitations, the interview technique can be used to yield useful information on how children solve problems. And, hopefully, teachers can do a better job of teaching children to solve problems once they better understand the problem-solving process.

BIBLIOGRAPHY

1. Anderson, R. C. (1964). Shaping logical behavior in six- and seven-year olds. Cooperative Research Project No. 1790A. Research Report to the U. S. Office of Education, Urbana, Illinois: University of Illinois.

2. Anderson, R. C., and Ausubel, D. P. (eds.) (1965). Readings in the psychology of cognition. New York: Holt, Rinehart and Winston.

3. Ausubel, D. P. (1960). The use of advanced organizers in the learning and retention of meaningful verbal material. Journal of Educational Psychology, 51: 267-272.

4. Ausubel, D. P. (1963). Cognitive structure and the facilitation of meaningful verbal learning. Journal of Teacher Education, 14: 217-21.

5. Ausubel, D. P. (1965). Introduction. In Anderson, R. C., and Ausubel, D. P. (eds.) Readings in the psychology of cognition. New York: Holt, Rinehart and Winston.

6. Berlyne, D. E. (1957). Recent developments in Piaget's work. British Journal of Educational Psychology, 27: 1-12.

7. Berlyne, D. E. (1960). Conflict, arousal and curiosity. New York: McGraw-Hill.

8. Bloom, B. S. (ed.) (1956). Taxonomy of educational objectives. Handbook I: Cognitive domain. New York: McKay.

9. Bloom, B. S., and Broder, Lois (1950). Problem-solving processes of college students. Supplementary Educational Monographs, No. 73. Chicago: The University of Chicago Press.

10. Bruner, J. S. (1961). The act of discovery. Harvard Educational Review, 31: 21-32.

11. Bruner, J. S. (1964). The course of cognitive growth. American Psychologists, 19: 1-15.

12. Bruner, J. S., Goodnow, J. J., and Austin, G. A. (1956). A study of thinking. New York: Wiley.

13. Buswell, G. T., and Kersh, B. Y. (1956). Patterns of thinking in solving problems. Univ. Calif. Publ. Educ., 12, No. 2.

14. Case, D., and Collinson, J. M. (1962). The development of formal thinking in verbal comprehension. British Journal of Educational Psychology, 32: 103-11.

15. Dewey, J. (1910). How we think. Boston: Heath.

16. Gagné, R. M. (1962). The acquisition of knowledge. *Psychological Review*, 69: 355-56.

17. Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart and Winston.

18. Guilford, J. P. (1965). *Fundamental statistics in psychology and education*, 4th ed. New York: McGraw-Hill.

19. Goss, A. E. (1961). Verbal mediating response and concept formation. *Psychological Review*, 68: 248-74.

20. Guilford, J. P. (1959). Three faces of intellect. *American Psychologists*, 14: 469-79.

21. Harper, R. J. C., Anderson, C. C., Christenson, C. M. and Hunka, S. M. (eds.) (1964). *The cognitive processes: readings*: Englewood Cliffs, N. J.: Prentice, Hall.

22. Hovland, C. I. (1960). Computer simulation of thinking. *American Psychologist*, 15: 687-93.

23. Hunt, J. McV. (1961). *Intelligence and experience*. New York: Ronald Press.

24. Johnson, D. M. (1955). *The psychology of thought and judgment*. New York: Harper.

25. Judson, A. J., Cofer, C. N., and Gelfand, S. (1956). Reasoning as an associative process: II. "Direction" in problem solving as a function of prior reinforcement of relevant responses. *Psychological Reports*, 2: 501-07.

26. Kendler, H. H., and Kendler, Tracy S. (1962). Vertical and horizontal processes in problem-solving. *Psychological Review*, 69: 1-16.

27. Loree, M. R. (1948). A study of a technique for improving tests. Unpublished doctor's dissertation, The University of Chicago.

28. Loree, M. R. (1965). *Psychology of education*. New York: Ronald Press.

29. Lovell, K. (1961). A follow-up study of Inhelder and Piaget's "The growth of logical thinking." *British Journal of Psychology*, 52: 143-53.

30. Miller, G. A., Galanter, E., and Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart and Winston.

31. Newell, A., Shaw, J. C., and Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, 65: 151-66.

32. Osgood, C. E. (1953). *Method and theory in experimental psychology*. Fair Lawn, N. J.: Oxford University Press.

33. Peake, Ronald (1964). An examination of the differences in problem-solving processes of successful and nonsuccessful students taking ninth grade social studies. Unpublished doctor's dissertation, University of Alabama.

34. Peel, E. S. (1959). Experimental examination of some of Piaget's schemata concerning children's perception and thinking and a discussion of their educational significance. British Journal of Educational Psychology, 29: 89-103.

35. Piaget, J. (1947). The psychology of intelligence. Trans. M. Piercy and D. E. Berlyne. London: Routledge and Kegan Paul, 1950.

36. Possien, Wilma M. (1964). A comparison of the effects of three teaching methodologies on the development of the problem-solving skills of sixth grade children. Unpublished doctor's dissertation, University of Alabama.

37. Rimoldi, H. J. A., et. al. (1962). Training in problem solving. Cooperative Research Project No. 1449. Research report to the U. S. Office of Education. Chicago: Loyola Psychometric Laboratory, Loyola University.

38. Rimoldi, et. al. (1964). Problem solving in high school and college students. Cooperative Research Project No. 2199. Research report to the U. S. Office of Education. Chicago: Loyola Psychometric Laboratory, Loyola University.

39. Roberts, John (1964). A study of the problem-solving processes of successful and nonsuccessful problem solvers in ninth grade science. Unpublished doctor's dissertation, University of Alabama.

40. Rokeach, M. (1960). The open and closed mind. New York: Basic Books.

41. Simon, H. A., and Newell, A. (1962). Computer simulation of human thinking and problem solving. Monograph of the Society for Research in Child Development, 27, No. 2: 137-50.

42. Staats, A. W., and Staats, Carolyn K. (1963). Complex human behavior. New York: Holt, Rinehart and Winston.

43. Thorndike, E. L. (1911). Animal intelligence. New York: Macmillan.

44. Travers, R. M. W. (1963). Essentials of learning. New York: Macmillan.

45. Vinacke, W. E. (1952). The psychology of thinking. New York: McGraw-Hill.

46. West, Charles K. (1965). Selectivity, redundancy, and contiguity as factors which influence the difficulty of an achievement test. Unpublished doctor's dissertation, University of Alabama.

SOCIAL STUDIES TEST ITEMS, INTERVIEW GUIDE AND SCORING PROCEDURES

APPENDIX I

SOCIAL STUDIES STEP FORM 3B PART I ITEMS 8 - 11
(Used for subjects grades 4 - 9)

A. TEST ITEMS

People act as they do for many different reasons. Here are four different reasons which may cause them to act as they do.

I. Because of the law of the place in which they live.
II. Because their families, friends, and neighbors act the same way.
III. Because it would not be good for their health if they acted otherwise.
IV. Because of the natural environment in which they live.

8. Which reason best explains why a great number of Americans like to attend baseball games?
   E I F II
   G III H IV

9. Which reason or pair of reasons best explains why people who live on sea coasts often earn their living by fishing?
   A I only B III only
   C II and III D II and IV

10. Which reason best explains why most American workers pay income taxes?
   E I F II
   G III H IV

11. Which pair of reasons best explain why people everywhere live in some kind of shelter?
   A I and II B I and III
   C II and IV D III and IV

NOTE: Flaps were placed over the four different reasons, and each of the four questions. The interviewer raises a flap over the four reasons or a question as requested by subject. Interviewer notes sequence as subject looks back and forth under flaps.

B. INTERVIEW GUIDE

Free-response portion
Subject is asked questions 8 through 11 in a free-response form. Thus:

8. Why do Americans like to attend baseball games?
9. Why do people who live on the sea coast often earn their living by fishing?
10. Why do most American workers pay income taxes?
11. Why do people everywhere live in some kind of shelter?

SOCIAL STUDIES STEP FORM 3B PART I ITEMS 8 - 11

B. INTERVIEW GUIDE (Continued)

Non-probing portion

Subject begins the non-probing portion of the interview by looking at the four reasons listed and the same questions as given in the free-response portion but in a multiple-choice form. Subject then reads aloud and answers each question 8 through 11. A flap covers up the reasons while the subject is reading a question. Another flap covers up the question while the subject is reviewing the reasons. The subject can request the interviewer to raise the "reasons" or "question" flap whenever he wishes. Interviewer records the sequence of requests of the subject and if subject does not verbalize his reason for requesting a certain flap to be raised (when he asks to look at it a second or third time) the interviewer asks him why he needs to see the material under the flap again.

Probing portion

1. For each question the interviewer asks the subject why he selected the answer that he did.
2. Interviewer makes sure that the subject understands the question by asking him to restate the problem in his own words.
3. If problem is not understood, interviewer restates the problem for the subject. All interviewers used the same or standard restatements of problems.
4. If subject fails to get the question correct even after understanding it, interviewer asks the subject if he would like some information. Interviewer continues to supply additional cues until the subject gets the question correct.
5. Suggested restatements of the problem to be provided if necessary:
    8. Most Americans attend baseball games because.....
    9. People who live on sea coasts earn their living by fishing because.....
    10. Most American workers pay income taxes because.....
    11. People everywhere live in some kind of shelter because.....

C. SCORING PROCEDURE

A rating of 1-5 to be given for each question on each of the four categories--(1) possession of information; (2) Retrieval of information; (3) extracting information from statement of problem; (4) combining operations.

SOCIAL STUDIES STEP FORM 3B PART I ITEMS 8 - 11

C. SCORING PROCEDURE (Continued)

1. Possession of information
   - (a) During any portion of the interview subject shows that he is able to supply a reason why people act as they do (i.e., attend baseball games, earn their living by fishing, etc.)
   - (b) Reveals an understanding of terms such as "natural environment," "shelter," "taxes," during any portion of the interview.

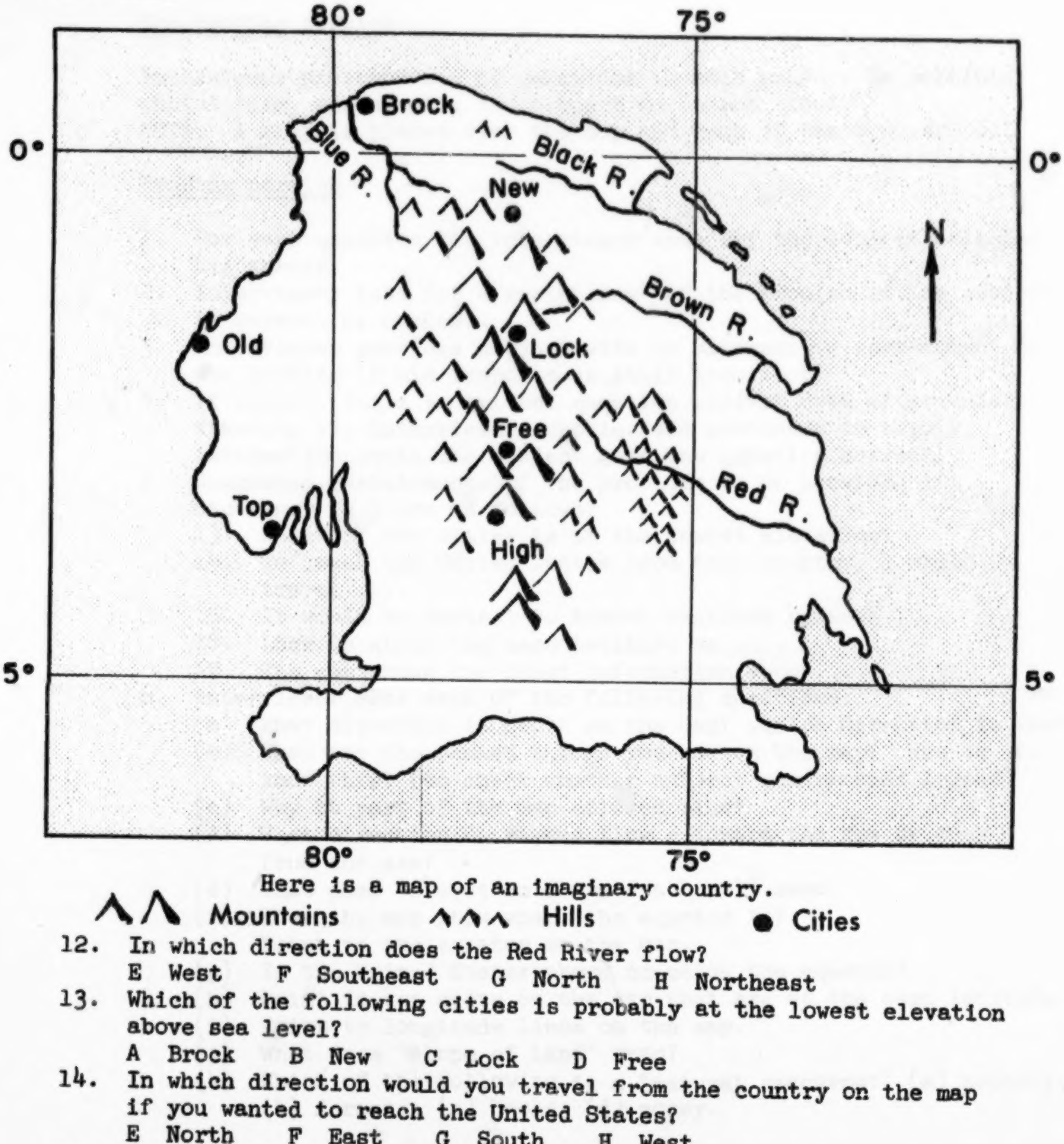
2. Retrieval of information
   - (a) In free-response portion of interview subject is able to supply reasons to questions "Why do Americans like to attend baseball games?" "Why do most Americans pay income taxes?" etc.
   - (b) In free-response portion of interview subject demonstrates understanding of key terms--"natural environment," "shelter," etc.

3. Extracting information
   - (a) Evidences understanding of what the question asks in the free-response form.
   - (b) For the multiple-choice form, understands that the answer is to be found among the alternative reasons I, II, III, IV.
   - (c) Evidences sufficient understanding of each of the four alternatives I, II, III, IV to proceed toward a solution of the problem.

4. Combining operations
   - (a) Searching behavior. Considers (looks over and evaluates) alternatives to find an answer to the question posed.
   - (b) Evidences reasoning in the process of selecting an answer.
   - (c) Is consistent. Chooses alternatives in probing portion consistent to those chosen in free response form.

SOCIAL STUDIES STEP FORM 3B PART I ITEMS 12 - 17
(Used for subjects grades 4-9)

A. TEST ITEM



Here is a map of an imaginary country.

ΛΛ Mountains ∧∧∧ Hills ● Cities

12. In which direction does the Red River flow?
E West F Southeast G North H Northeast

13. Which of the following cities is probably at the lowest elevation above sea level?
A Brock B New C Lock D Free

14. In which direction would you travel from the country on the map if you wanted to reach the United States?
E North F East G South H West

15. Between which of the following cities would overland travel be easiest?
A Brock and Lock B Top and High C Old and Free D Top and Old

16. Lock is located at about the same latitude as the city of
E New F Old G Brock H Top

17. The map provides the LEAST information concerning which of the following factors?
A Longitude B Comparative distance
C Slope of land D Natural resources

SOCIAL STUDIES STEP FORM 3B PART I ITEMS 12 - 17 (Continued)

B. INTERVIEW GUIDE

Non-probing portion

Interviewer presents set of questions to each subject in multiple-choice form, encouraging the subject to "think aloud."
NOTE: A flap is placed over the map and each of the 6 questions.

Probing portion

1. For each question the interviewer asks why the subject selected his answer.
2. Interviewer asks for a restatement of the problem if the subject is correct or unclear.
3. Interviewer provides subject with an alternative restatement of the problem if his response is still incorrect.
4. If subject fails to get the question correct even after understanding it, interviewer supplies and continues to supply information until the subject gets the question correct.
5. Suggested restatements of the problems to be provided if necessary in 3 are as follows:
   13. Which of the cities is at the lowest elevation?
   14. To reach the United States from this country, I would travel.....
   15. It would be easiest to travel overland between.....
   16. Lock is about the same latitude as.....
   17. The map shows the least information about.....
6. Interviewer asks each of the following questions:
   (a) What direction is north on the map? Which direction is east?
   (b) What are the peaked things you see on the map? How do you know this? (to check whether subject understands legend)
   (c) Why is part of the map colored blue?
   (d) What direction do rivers flow - toward the sea or away from the sea?
   (e) What does "elevation above sea level" mean?
   (f) Does the map show where the equator is?
   Point to the equator on the map.
   (g) Is the United States above or below the equator?
   (h) Point to two spots on the map that are at the same latitude.
   (i) Point to longitude lines on the map.
   (j) What does "slope of land" mean?
   (k) Which of the following is a "natural resource?" (a) honesty; (b) forests; (c) books; (d) money.

Card Sort Portion

"Now let us look at Questions 12 through 17 again....I have here 3 cards."

Card 1 reads: South is toward the bottom of the map; west is toward the left on the map.
Card 2 reads: Rivers flow toward the sea; not away from the sea.

SOCIAL STUDIES STEP FORM 3B PART I ITEMS 12 - 17

B. INTERVIEW GUIDE

Card Sort Portion (Continued)

Card 3 reads: The Red River is south of the Brown River.

(Ask or discover about each of the following questions:)

1. Which card or cards contain information that is needed to answer question 12? (More than one card may be needed)
2. Tell me why you made that selection. (For this task make sure the subject understands what you are asking him to do. Explain why Cards 1 and 2 are needed)

"Now let us look at Question 13. Here are three different cards:"

Card 1 reads: Free is south of New.
Card 2 reads: Cities at the seacoast are of lower elevation than cities in the mountains.
Card 3 reads: Lock is located near two rivers.
"Which card or cards contain information that is needed to answer question 13?" "Why?"

"Now let us look at Question 14. Here again are three cards:"

Card 1 reads: The central part of the imaginary country contains many mountains.
Card 2 reads: Most of the imaginary country lies south of the equator.
Card 3 reads: The United States is north of the equator.
"Which card or cards contain information that is needed to answer question 14?" "Why?"

"Now let us look at Question 15. Here are three cards:"

Card 1 reads: Traveling over mountains is not easy.
Card 2 reads: Top is west of High.
Card 3 reads: Brock is close to the Blue River.
"Which card or cards contain information that is needed to answer Question 15?" "Why?"

"Now let us look at Question 16. Here are three cards:"

Card 1 reads: Most of the imaginary country lies between 80° and 75° longitude.
Card 2 reads: Places that are equal distances from the equator are at the same latitude.
Card 3 reads: The Black River, Brown River, and Red River are at different latitudes.
"Which card or cards contain information that is needed to answer question 16?" "Why?"
No card sort for Question 17

C. SCORING PROCEDURE

Rating of 1-5 for each question on each of the four categories.

Question 12

1. Possession of information
   (a) Knows how to find east given direction north
   (b) Knows that rivers flow toward the sea

2. Retrieval of information
   (a) Recalls that rivers flow toward the sea
   (b) Recalls how to find directions

3. Extracting information
   (a) Understands that he is to find the direction that the Red River flows
   (b) Searches
   (c) Locates the Red River on the map
   (d) Ignores irrelevant information

4. Combining operations
   Solves the problem by combining 3(a), 3(b), 2(a) and 2(b).

Question 13

1. Possession of information
   (a) Understands the term "lowest elevation above sea level"

2. Retrieval of information
   (a) Recalls that places on seacoast are at a "low elevation above sea level" and that
   (b) Places in mountains are "high above sea level"

3. Extracting information
   (a) Understands question asked
   (b) Searches
   (c) Notes that Brock is on the seacoast or that the remaining cities are in or close to the mountains
   (d) Ignores the irrelevant on the card sort

4. Combining operations
   Solves problem by combining 3(a) and 3(b)

C. SCORING PROCEDURE (Continued)

Question 14

1. Possession of information
    (a) Knows that the United States is above, or north of the equator
    (b) Knows that the equator is at 0° latitude

2. Retrieval of information
    (a) Evidences use of 1(a) and 1(b) in solving problem
    (b) Chooses cards 2 and 3 in card sort

3. Extracting information
    (a) Understands the question
    (b) Searches
    (c) Identifies equator on the map

4. Combining operations:
    (a) United States is north of the Equator
    (b) The map of the imaginary country is at the south of the equator
    (c) Therefore one must travel north to reach the United States from the imaginary country

Question 15

1. Possession of information
    (a) It is difficult to travel over mountains

2. Retrieval of information
    (a) Evidences use of 1(a) in answering question

3. Extracting information
    (a) Understands question
    (b) Searches
    (c) Elimination of alternatives A, B, C on basis of 1(a)

4. Combining operations
    Solving problem by comparison of A, B, C and D and rejection of A, B, C, on basis of mountain travel, or solves on basis of lack of mountains between Top and Old

Question 16

1. Possession of information
    (a) Knows meaning of "latitude"
    (b) Chooses only card 2 in card sort

C. SCORING PROCEDURE

Question 16 (Continued)

2. Retrieval of information
   Evidences use of 1(a) in answering question (not later)

3. Extracting information
   (a) Understands problem
   (b) Searches generally
   (c) Locates Lock on map
   (d) Searches specifically for city that is of same latitude

4. Combining operations
   Solves problem logically by combining 1(a) and 1(b); 3(a), 3(b), 3(c) and 3(d).

Question 17

1. Possession of information
   (a) Meaning of longitude
   (b) Evidences understanding of "slope of land"
   (c) Meaning of natural resources

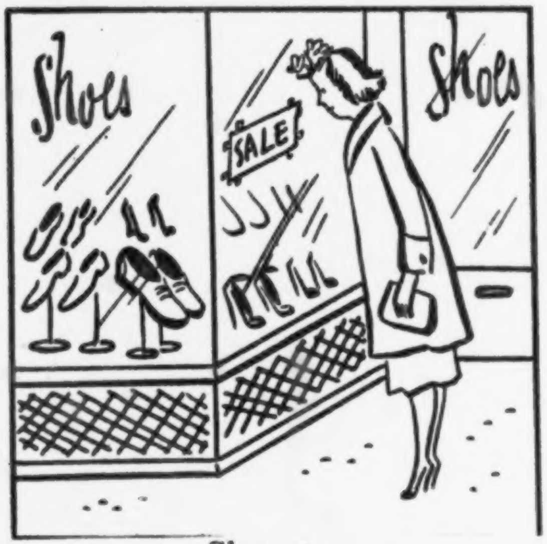
2. Retrieval of information
   (a) Knows meaning of longitude
   (b) Mountain symbols provide information about slope
   (c) Knows that forest are a natural resource

3. Extracting information
   (a) Understands the question
   (b) Searches
   (c) The meaning attached to longitude lines, mountain symbols, and location of cities
   (d) Does not miss question because of "natural resources" (That is, if subject answered correctly in natural resource probe of question and missed question 17)

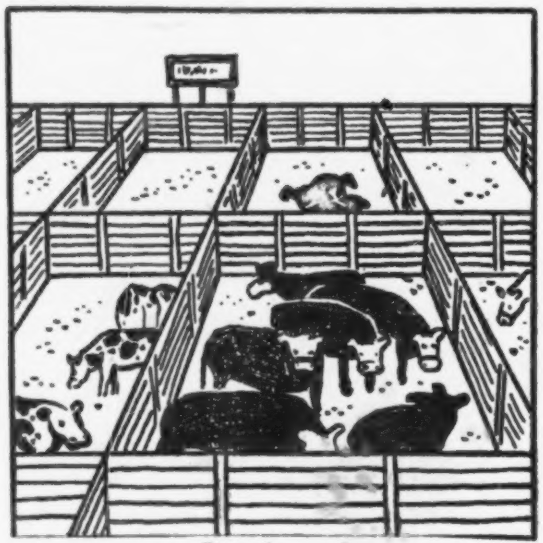
4. Combining operations
   Using 3 above as a basis for eliminating A, B, and C.

SOCIAL STUDIES STEP FORM 4B PART I ITEMS 21 - 25

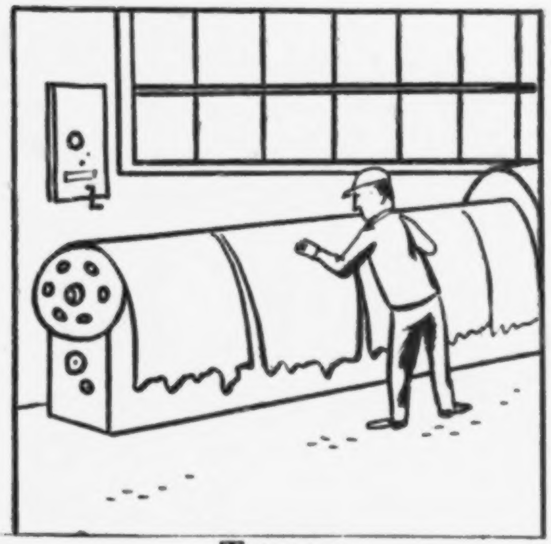
A. TEST ITEMS (Used for subjects grades 4-6)
These pictures tell the story of shoes. They are not in the right order.



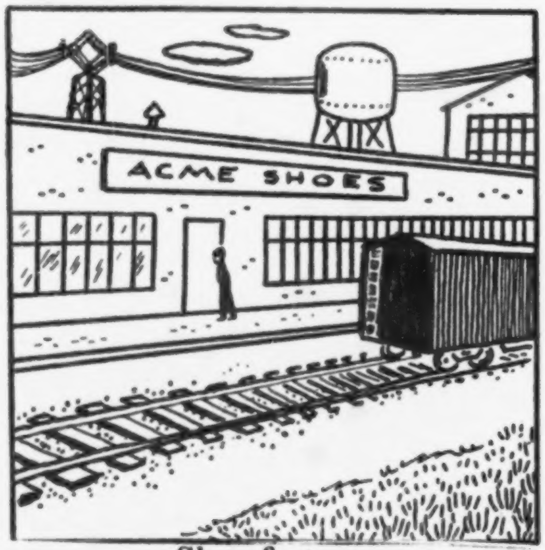
Shoe store



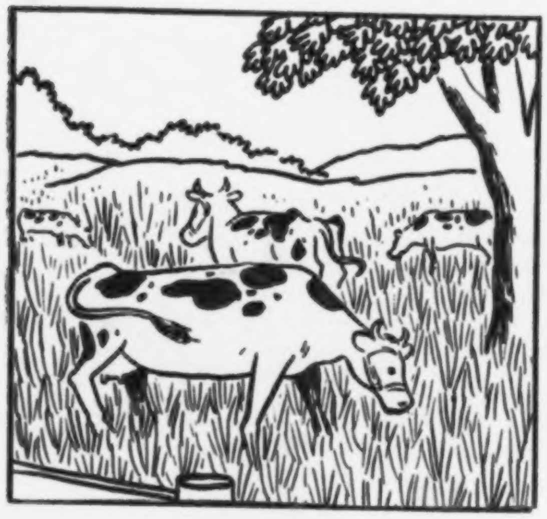
Stockyard



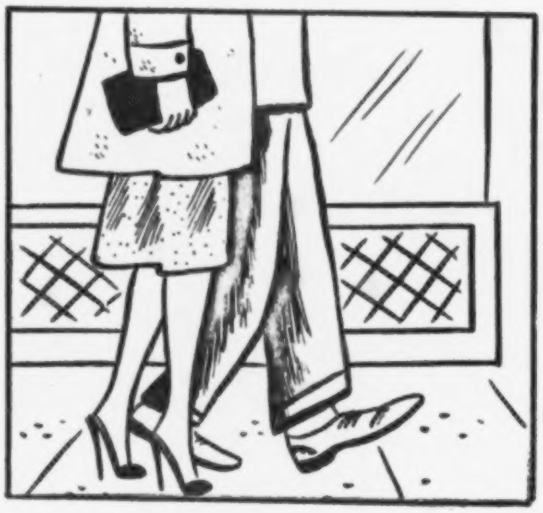
Tannery



Shoe factory



Pasture



Shoes in use

SOCIAL STUDIES STEP FORM 4B PART I ITEMS 21 - 25

A. TEST ITEMS (Continued)

21. Which picture should come first?
    A The stockyard
    B The tannery
    C The shoe factory
    D The pasture

22. Which picture should come next after the tannery?
    E The stockyard
    F The shoe factory
    G The pasture
    H The shoe store

23. In which of the pictures does electricity play the most important part?
    A The shoe store
    B The stockyard
    C The shoe factory
    D The pasture

24. Which of the pictures about shoes could also be shown in the story of meat?
    E The stockyard and the tannery
    F The tannery and the pasture
    G The shoe factory and the pasture
    H The stockyard and the pasture

25. Which of the pictures shows something that most people see most often?
    A The shoe store
    B The stockyard
    C The tannery
    C The shoe factory

NOTE: One flap is placed over the set of six pictures about shoes, and flaps are placed over each of the five questions.

B. INTERVIEW GUIDE

Non-probing portion

Interviewer presents set of questions to each subject in multiple-choice form, encouraging the subject to "think aloud."

Probing portion

1. Interviewer asks the following questions:
    (a) What are the shoes that you see in the store window made of?
    (b) What kind of animals do you see in the stockyard?
    (c) What work is done in a tannery?
    (d) Why are the cows penned up in the stockyard?
    (e) What is going to happen to the cows?
    (f) Why do you think the man is entering the factory?
    (g) What is the man in the tannery picture looking at?
    (h) Do most people buy their shoes at a shoe factory?
    (i) Would you expect to find any large machines at a shoe factory?

2. For each test item ask the following questions
    (a) For question number __ you said that __ answer was correct. Why did you think so?
    (b) Why not ___ (an alternate answer)?
    (c) Why not ___ (an alternate answer)?
    (d) Why not ___ (an alternate answer)?

SOCIAL STUDIES STEP FORM 4B PART I ITEMS 21 - 25

B. INTERVIEW GUIDE

Probing Portion (Continued)

3. Give subject an opportunity to revise his answer if he so desires.

C. SCORING PROCEDURES
Rating of 1-5 for each question on each of the four categories.

Question 21

1. Possession of information
   (a) Has some idea of the sequence of operations by which cattle become shoes, evidenced anywhere in the protocol.

2. Retrieval of information
   (a) Recalls in the initial portion of the interview that the start of the story of shoes takes place in a pasture.

3. Extracting information
   (a) Evidences a strategy that leads to an understanding of the question posed. Partial credit can be given (2 or 3) if strategy develops on the probing portion.

4. Combining operations
   Solves problem by combining his knowledge, his understanding of the question and the requirements of the question.

Question 22

1. Possession of information
   (a) Knows what work is done in a tannery
   (b) Gives evidence anywhere in the interview of having some idea of what work is done in a shoe factory

2. Retrieval of information
   (a) Draws upon his knowledge of sequence of activities and what happens in a tannery and a shoe factory. Score less when retrieval occurs in probe.

3. Extracting information
   (a) Evidences understanding of question.

4. Combining operations
   Combines: (a) Memory of sequence
   (b) Understanding of problems
   (c) Information in pictures to arrive at answer

C. SCORING PROCEDURES (Continued)

Question 23

1. Possession of information
   (a) Knows that large machines would be required in a shoe factory with electricity to operate them.

2. Retrieval of information
   (a) Recalls 1(a) in probe and gives answer consistent to this recalled information in the unstructured portion

3. Extracting information
   (a) Understands the question

4. Combining operations
   (a) Combines 3(a) and 2(a)

Question 24

1. Possession of information
   (a) Knows that cows slaughtered in the stockyard yield several products - meat and hides
   (b) Knows that cows go from pasture to the stockyard (Give credit anywhere in protocol)

2. Retrieval of information
   (a) Recalls the first facts (pasture and stockyard) in the story of meat

3. Extracting information
   (a) Understands the question (Give credit unless child clearly doesn't)

4. Combining operations
   (a) Combines 2(a) and 3(a)
   (b) Gives good reason for not accepting other alternatives

Question 25

1. Possession of information
   (a) Knows that many people see shoe stores (Give credit anywhere)

2. Retrieval of information
   (a) Recalls 1(a)

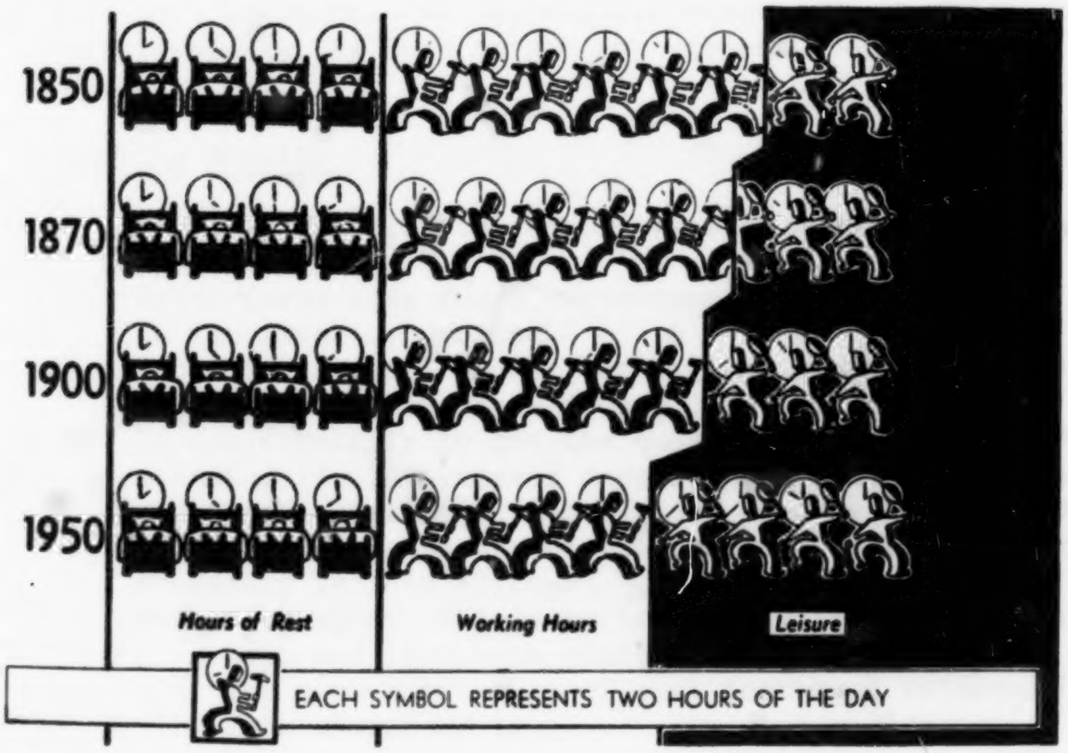
3. Extracting information
   (a) Understands the question

4. Combining information
   (a) Combines 2(a), 3(a) and 3(b)

SOCIAL STUDIES STEP FORM 3B PART II ITEMS 24 - 28

A. TEST ITEMS (Used for subjects grades 7-9)

The chart below shows the changing amounts of time the average American spent in leisure and work between 1850 and 1950.



Below are several statements. You are to decide whether each statement is

I. *true* and the chart proves that it is true,
II. *true but* you would need more information than the chart gives you to prove that it is true,
III. *false* and the chart proves that it is false,
IV. *false but* you would need more information than the chart gives you to prove that it is false.

24 The proportion of time spent in leisure time activities has increased between 1850 and 1950. This statement is

E I **F** II **G** III **H** IV

25 The average number of hours of rest increased between 1850 and 1950. This statement is

A I **B** II **C** III **D** IV

26 The number of people working in 1850 was larger than the number of people working in 1950. This statement is

E I **F** II **G** III **H** IV

27 The average American worked two hours more a day in 1850 than in 1950. This statement is

A I **B** II **C** III **D** IV

28 Farmers have a larger amount of leisure time now than they formerly had. This statement is

E I **F** II **G** III **H** IV

SOCIAL STUDIES STEP FORM 3B PART II ITEMS 24 - 28 (Continued)

B. INTERVIEW GUIDE

Non-probing portion

Flaps were placed over (1) the chart, (2) the answer key, and (3) each question. As the subject reads each question and gives his responses, be sure to get why subject goes to a flap for a second time.

Probing Portion

1. For each question the interviewer asks why the subject selected the answer he did.
2. Interviewer asks for a restatement of the problem if the subject is incorrect or unclear.
3. Interviewer provides subject with an alternative restatement of the problem if his response is still incorrect.
4. If subject fails to get the question correct even after understanding it, interviewer supplies and continues to supply additional cues until the subject does get the question correct.
5. Suggested restatements of the problems to be provided if necessary in 3 as follows:
    24. Time spent in leisure time activities has increased between 1850 and 1950 this statement is.....
    25. The number of hours of rest increased between 1850 and 1950. This statement is.....
    26. No change
    27. No change
    28. Farmers have more leisure time than they used to have. This statement is.....
6. If subject is unsuccessful interviewer probes further to find out if subject understands key terms and meanings of symbols.

Card Sort

INTERVIEWER:
"I have here some cards. Each card has some information. You are to sort the cards into two piles. If the statement on the card is also contained on the chart, place the card in Pile A; if the statement on the card is not contained on the chart, place the card in Pile B, I will start you off. On this card it says: "..."(1) Is this information contained on the chart?" (Give subject a chance to answer) "No, it isn't, so we will place this card in Pile B. On this card it says: "..."(3) Is this information contained in the chart?" (Give subject a chance to answer) "Yes, it is contained in the chart. Now read each of these cards aloud and tell me in which pile it should be placed. Remember if the information on the card is contained on the chart it goes in Pile A, if not in Pile B."

SOCIAL STUDIES STEP FORM 3B PART II ITEMS 24 - 28

B. INTERVIEW GUIDE (Continued)

Card Sort

Card 1. In 1900 the average American worked 7 hours.
Card 2. There were more types of leisure activity in 1950 than in 1850.
Card 3. In 1950, the average American spent 8 hours a day in leisure time activities.
Card 4. In 1850, the average American spent about 8 hours in leisure time activities.
Card 5. In 1950, many Americans spent some of their leisure time playing golf.
Card 6. Many Americans were farmers in 1850.
Card 7. In 1850, the average American spent 8 hours a day resting.
Card 8. Many people owned automobiles in 1850.
Card 9. In 1850 people used horses to get to work.
Card 10. In 1950 the average American rested 8 hours a day.

INTERVIEWER:

"Now you have made your two piles. I am going to rearrange these cards a bit." (Interviewer corrects mistakes of subject by putting in Pile A only the cards belonging in Pile A. Interviewer jots down the cards sorted incorrectly before proceeding).
"Now I want you to look through these cards (Cards: 2, 3, 4, 7, 10) and select only the cards that contain the information you would need to answer problem 24." (Put cards back in pile before proceeding to problem 25). "Now look through the cards and select the cards containing information you would need to answer problem 25."

C. SCORING PROCEDURE

Rating of 1-5 for each question on each of the four categories.

Question 24

1. Possession of information
    - (a) Meaning of "proportion of time"
    - (b) Meaning of "symbol"
    - (c) Meaning of "leisure time"

2. Retrieval of information
    - (a) Retrieves 1(a), (b), (c)

3. Extracting information
    - (a) Understands the general nature of the problem series
    - (b) Able to read the chart and understand the different symbols
    - (c) Selects card 3 on second card sort

SOCIAL STUDIES STEP FORM 3B PART II ITEMS 24 - 28

C. SCORING PROCEDURES

Question 24 (Continued)

4. Combining operations
   (a) Seeks answer to question in chart
   (b) Considers "leisure time" portion of chart
   (c) Reasons in selection of answer (Comparisons, e.g., 1850 to 1950) Evaluates, e.g., in terms of true-false key (I, II, III, IV)
   (d) Selects card 3 on second card sort

Question 25

1. Possession of information
   (a) Meaning of "hours of rest"

2. Retrieval of information
   (a) Retrieves 1(a)

3. Extracting information
   (a) Understands the general nature of the problem series
   (b) Able to read the chart and the various symbols
   (c) Selects card 7 on first card sort

4. Combining operations
   (a) Seeks answer to question in chart
   (b) Focuses on "rest" portion of chart
   (c) Reasons in selection of answer (Comparisons and evaluations - see Question 24, 4(c))
   (d) Selects card 7 and 10 on second card sort

Question 26

1. Possession of information
   (a) More people worked in 1950 than in 1850

2. Retrieval of information
   (a) Retrieves 1(a)

3. Extracting information
   (a) Understands the general nature of the problem series
   (b) Able to read the chart and the various symbols
   (c) Translates symbols as amount of time NOT number of people
   (d) Does not select 1, 2, or 4 on first card sort

C. SCORING PROCEDURES

Question 26 (Continued)

4. Combining operations
    (a) Seeks answer to question in chart
    (b) Notes that chart does not tell the number of people working in 1950 or 1850
    (c) Uses fact that more people worked in 1950 than 1850 in selecting answer
    (d) Reasoning (Comparisons and evaluations, See 24, 4(c))
    (e) Rejects all cards on second card sort

Question 27

1. Possession of information
    (a) ____________________ Give 3 points unless you can detect a lack of knowledge

2. Retrieval of information
    (a) ____________________ Give 3 points unless you can detect a lack of retrieval

3. Extracting information
    (a) Understands the general nature of the problem series
    (b) Able to read chart and the various symbols
    (c) Notes symbols represent 2 hours of the day
    (d) Does not select 5, 6, 8, and 9 on first card sort

4. Combining operations
    (a) Seeks answer on chart
    (b) Focuses on working hours section of chart
    (c) Reasons (Comparisons and evaluations, See 24, 4(c))
    (d) Rejects all cards on second card sort

Question 28

1. Possession of knowledge
    (a) Has general idea that leisure time has increased for most people

2. Retrieval of information
    (a) Retrieves 1(a)

C. SCORING PROCEDURES

Question 28 (Continued)

3. Extracting information
    (a) Understands the general nature of the problem series
    (b) Able to read chart and various symbols
    (c) Notes that chart applies to average American, not farmer
    (d) Selects card 10 on first card sort

4. Combining operations
    (a) Seeks answer in chart
    (b) Focuses on "leisure time" activities
    (c) Makes judgment concerning farmers leisure time
    (d) Reasoning (Comparisons and evaluations, See 24 4(c))
    (e) Rejects all cards on card sort

SCIENCE TEST ITEMS, INTERVIEW GUIDE
AND SCORING PROCEDURES

APPENDIX II

SCIENCE STEP FORM 4B PART I ITEM 12
(Used for subjects grades 4-9)

A. TEST ITEM

Toy Shop

One day some boys and girls went to the toy shop to buy and look at toys.

12. Mary and Jane each bought the same kind of rubber ball in the store. They went out of the store and played with them. Mary said, "My ball bounces higher than yours, Jane." Jane replied, "I'd like to see you prove that." What should Mary do?
    - E Drop both balls at the same time from the same height and measure how far each one bounces.
    - F Throw both balls against a wall and see how far each ball bounces off the wall.
    - G Drop each ball from a different height at the same time and measure the height of each bounce.
    - H Throw the balls down against the floor and see how high they bounce.

B. INTERVIEW GUIDE

Non-probing portion

Present question with responses

Probing portion

1. Can you tell me what the question asks you to do?
2. Why did you select answer (--)?
3. Why do balls bounce?
4. What are some possible reasons why one ball would bounce higher than another ball?
5. Does how hard you throw a ball have anything to do with how far it will bounce?
6. If we dropped the same ball from the same height twice in a row, would it bounce the same distance both times?
7. Why would dropping the ball from different heights fail to prove your experiment?
8. What effect would it have on the experiment if the same person couldn't throw the ball just as hard one time as another?
9. Can you think of a different way to do this experiment?

SCIENCE STEP FORM 4B PART I ITEM 12 (Continued)

C. SCORING PROCEDURES

1. Possession of information*
   Knows:
   - (a) Why balls bounce (3)
   - (b) Possible reasons why one ball bounces higher than another (4)
   - (c) Facts of use in assessing factors affecting how high a ball bounces:
     - -how hard you throw the ball (5, 8)
     - -the height from which you drop a ball (6, 7)

2. Retrieval of information
   Evidences retrieval of 1 (a), (b), (c) either in unstructured portion of interview or in response to the question "Why did you select answer ___?"

3. Extracting information
   - (a) Understands question

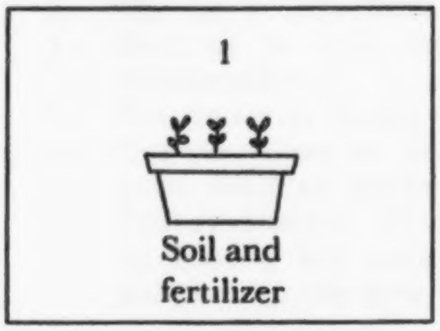
4. Combining operations
   Rating
   - 5-4 Evaluates experiments in terms of knowledge of science, and/or is able to plan an experiment with proper controls.
   - 3-2 Evaluates experiments in terms of personal experiences with bouncing balls, and is able to plan some kind of experiment even though faulty.
   - 1- Unable to evaluate experiments or plan experiment

*Numbers in parenthesis refer to the number of corresponding question in the Interview Guide

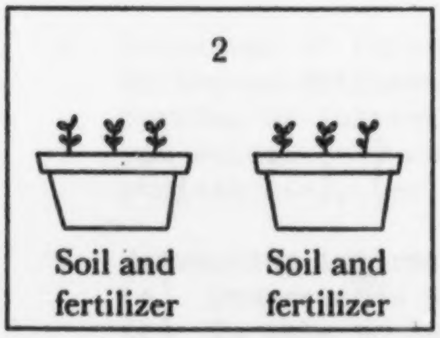
SCIENCE STEP FORM 4B PART I ITEM 19
(Used for subjects grades 4-9)

A. TEST ITEM

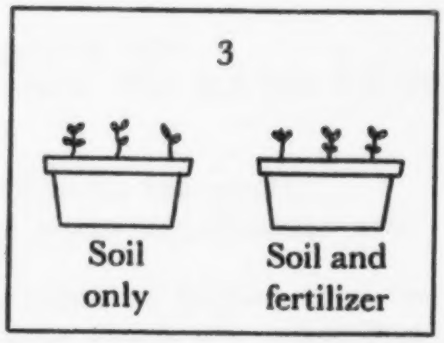
19 Sue wanted to test some of her father's fertilizer to see if it would really make plants grow better. She thought of three different experiments.


1

Experiment 1. Add fertilizer to one flowerpot of soil. Plant three bean seeds. Watch them grow while giving them each the same care.


2

Experiment 2. Get two flowerpots. Add fertilizer to the soil of both pots. Plant three bean seeds in each pot. Watch them grow while giving them each the same care.


3

Experiment 3. Get two flowerpots. Add fertilizer to one flowerpot. Plant three bean seeds in each. Watch them grow while giving them the same care.

Which experiment would show if this fertilizer was good for their plants?

A Experiment 1
B Experiment 2
C Experiment 3
D None of the experiments above will tell her anything about the fertilizer.

SCIENCE STEP FORM 4B PART I ITEM 19 (Continued)

B. INTERVIEW GUIDE

Non-probing portion

Present problem with responses

Probing portion

1. What is it that the question asks you to do?
2. Why did you select (--) as your answer? Why did you not select choices (--), (--), (--)?
3. What is an experiment?
4. Why do people sometimes add fertilizer to the soil?
5. Tell me in your own words how Sue's three experiments are different?
6. How can you judge whether one experiment is better than another?
7. Can you name or describe some products which use experiments like this in their advertising? See how many you can name. (Interviewer: Give subject about 15 seconds...if successful in naming any examples, give clue of Crest. But, if examples subject gives are merely incorrect, let him continue on.)

C. SCORING PROCEDURES

1. Possession of information
   (a) Able to explain fairly adequately what an experiment is (3)
   (b) Knows the use of fertilizer (4)
   (c) Able to supply some criteria for a good experiment (6)

2. Retrieval of information
   Evidences retrieval of 1(a), (b), (c) either in unstructured portion of interview or in response to the question "Why did you select (--) as your answer?" "Why did you not select choices (--), (--), and (--)?" (2)

3. Extracting information
   (a) Understands the question (1)
   (b) Is able to tell how Sue's experiments differ (5 or 2)

4. Combining operations

Rating		
	5-4	Evaluates Sue's experiments in terms of a knowledge of the nature of an experiment and the function of fertilizers
	3-2	Exhibits ability to evaluate Sue's experiments only after help from the interviewer
	1-	Unable to evaluate Sue's experiments

SCIENCE STEP FORM 4B PART I ITEMS 28 and 29
(Used for subjects grades 4-9)

A. TEST ITEMS

Taking A Rocket Trip To The Moon

Let's imagine you are taking a trip to the moon in a rocket ship.

28. The moon seems to get larger and larger as you get closer to it. Which one of the following reasons best explains this?
    - E The nearer you are to an object, the bigger it seems to you
    - F On the earth, the air makes the moon seem smaller than it should
    - G The atmosphere of the moon acts like a telescope
    - H Out in space, both the moon and the stars receive more light from the sun

29. The sun is the only object in our solar system that makes and gives off large amounts of heat and light. This means that we see the moon because it
    - A reflects light from the sun
    - B gives off it own light
    - C is not in our solar system
    - D reflects much light from other planets

B. INTERVIEW GUIDE

Question 28

Non-probing portion

(a) Present problem open-ended
(b) Present four responses

Probing portion

1. What does the question ask you to do?
2. Why did you select your answer? (INT. - do not probe or ask "Why not" for the other alternatives.)
3. Which of the following are instances of the same kind as the question?
   (a) A blown up balloon is bigger than one which is not blown up
   (b) Telephone poles in the distance look smaller than those close up
   (c) An airplane looks smaller in the air than it does on the ground
4. Here is a dime and here is a quarter: Which is further away from you?*

*A card is shown with pictures of a dime and a quarter. Both coins are drawn the same size.

SCIENCE STEP FORM 4B PART I ITEMS 28 and 29

B. INTERVIEW GUIDE (Continued)

5. Here is a line of telephone poles. This is a railroad track. What's wrong with this picture?**
6. Here is a picture of the same balls at different distances. What is wrong with this picture?***

**A card is shown picturing telephone poles along side a railroad track. The railroad tracks appear to become closer together in the distance. However, the telephone poles as they recede in the distance are drawn larger.

***Balls are drawn larger as they recede in the distance.

C. SCORING PROCEDURES

Question 28

1. Possession of information
   Rate 3 unless unusual instance shows how information has facilitated or hindered problem solution.

2. Retrieval of information
   Rate 3.

3. Extracting information
   (a) Understands the question (1)
   (b) Focuses attention on "The moon seems to get larger and larger as you get closer to it"

4. Combining operations
   (a) Recognition that the moon seeming to get larger is just one instance of the generalization: "The nearer you are to an object, the bigger it seems to you."
   (b) Ability to recognize other instances of the generalization and discriminate such instances from non-instances (3, 4, 5, 6)

B. INTERVIEW GUIDE

Question 29

Non-probing portion

Present problem with responses

Probing portion

1. Can you tell me what the problem asks you to do?

SCIENCE STEP FORM 4B PART I ITEMS 28 and 29

B. INTERVIEW GUIDE

Probing portion (Continued)

2. Why did you select the answer you did?
3. (INT. - Here we want to find out at what level of generality the subject is operating, and if he can identify the sun, earth and moon as being part of our solar system
   - (a) What is a solar system (INT. - If subject does not answer completely, ask:)
   - (b) What is our solar system? (INT. - If subject rambles, let him go on - do not probe.)
   - (c) Is the moon in our solar system?
4. What is the difference between a planet and a star?
5. Is the sun a planet or a star?
6. Do planets give off large amounts of heat and light?
7. Does the earth give off large amounts of heat and light? Why or why not?
8. Does the moon give off its own light or reflected light?
9. Where does the light in our solar system come from?
10. Why did you not select (--), (--), or (--)?

C. SCORING PROCEDURES

Question 29

1. Possession of information
   - (a) Knows what a solar system is, and what our solar system is (5), and that the moon is in our solar system (3)
   - (b) Knows the difference between a planet and a star (4)
   - (c) Knows that the sun is a star (5)

2. Retrieval of information
   - (a) Evidences recall of some of the information in 1(a), (b) or (c) in explaining why answer selected (2)

3. Extracting information
   - (a) Understands the question asked (1)
   - (b) Focuses attention on the statement "The sun is the only object in our solar system that makes and gives off large amounts of heat and light" in answering interview question 6, 7, 8, and 9 (2)

4. Combining operations
   - (a) Uses recalled information and information contained in question in answering question (2)
   - (b) After series of questions 3-9, has clarified the question and is able to answer interview question 10.

SCIENCE STEP FORM 4B PART II ITEMS 13 and 18
(Used for subjects grades 4-9)

A. TEST ITEMS

A Visit To A Ranch

One summer Ralph visited his uncle's ranch in Wyoming. One of the things he liked best was riding his uncle's horses.

13. As he was riding one day he found a little stream. He had heard his uncle say that it usually rains very little in Wyoming, so he wondered where the water came from. Which one of the following would be most likely to help answer his question?
    A Looking at the color of the water
    B Riding downstream until he meets a larger stream
    C Counting the days to the next rainstorm
    D Riding his horse upstream

18. The well on the ranch is almost dry and his uncle is going to have it dug deeper. He tells him that the water has been dropping lower in the well for the past several years. Which one of the following would LEAST help to explain this?
    E Too much water has been running off the land into the rivers
    F People have been pumping too much water from wells
    G There has been less rain
    H There has been too much snow falling on the mountain tops

B. INTERVIEW GUIDE

Question 13

Non-probing portion

Present problem 13 with responses.

Probing portion

1. What does the question ask you to do?
2. Which way does the water flow? (Tell me what you mean, give me an example, etc.)
3. How could Ralph find out which way the water was flowing?
4. Does water flow upstream or downstream?
5. (INT. - If subject is initially unable to answer "What does the question ask you to do," ask again using a map as a visual aid).
6. (INT. - For successful subject ask:) Why did you not select answer (--), (--), or (--)?

SCIENCE STEP FORM 4B PART II ITEMS 13 and 18

C. SCORING PROCEDURES

Question 13

1. Possession of information
   - (a) Water flows from a higher level to a lower level (2); or another correct response to question 2.
   - (b) Water flows downstream
   - (c) Water flows in a stream

2. Retrieval of information

   Rating
   - 5-4 Recalls 1(a) or 1(b) in unstructured portion of interview
   - 3-2 Recalls 1(a) or 1(b) during structured portion of interview
   - 1- Does not recall 1(a) or 1(b)

3. Extracting information
   - (a) Understands the question (1)
   - (b) Focuses on essential elements of test item - What would have to be done to find where the water came from?
   - (c) Is not distracted by useless information contained in test item

4. Combining operations

   Rating
   - 5-4 Applies information that water flows downstream in evaluating alternative responses
   - 3-2 Successful in interpreting question only in response to interview question 5
   - 1- Unsuccessful in answering question

B. INTERVIEW GUIDE

Question 18

Non-probing portion

Present problem 18 with responses

Probing portion

1. What does the question ask you to do?
2. (a) Is there anything which you have read or seen which tells you anything about water underground?
   - (b) Do E, F, G, or H (response choices) suggest any material you have read or seen about the amount of water underground?
3. When water runs off the land, how would this affect the amount of water in the well?

SCIENCE STEP FORM 4B PART II ITEMS 13 and 18

B. INTERVIEW GUIDE

Question 18

Probing portion (Continued)

4. When you pump water out of a well, what happens to the amount of water in the well?
5. What does the amount of rain have to do with the amount of water in a well?
6. (a) What effect does lots of snow falling on mountain tops have to do with the amount of water in a well?
   (b) Could some well water come from snow?
7. (INT. - For use with non-successful subject)
   (a) What do you think the question asks?
   (b) Would you select an answer again and give your reason for your answer?

C. SCORING PROCEDURES

Question 18

1. Possession of information
   (a) General science knowledge about underground water (2)
   (b) Water running off the land would affect the amount of water reaching the well (3)
   (c) Pumping water out of a well reduces the amount of water in the well (4)
   (d) Rain can add to the amount of water in the well (5)
   (e) Snow on mountain tops could eventually result in an increase of water (6a)

2. Retrieval of information
   Rating
   5-4 Evidences in unstructured portion of interview some recall of above knowledge
   3-2 Evidences of recall occurs in probing portion of interview
   1- No evidence of recall

3. Extracting information
   (a) Understands the question (1)
   (b) Does not misinterpret because of the word "least"

4. Combining operations
   Rating
   5-4 Evaluates alternatives in terms of their effect on water level in unstructured portion of interview or before interview question 7 is reached in structured portion
   3-2 Evaluates alternatives in interview question 7
   1- Failure to relate alternatives to water level of well

SCIENCE STEP FORM 4B PART II ITEMS 20, 22, 24, 25
(Used for subjects grades 4-9)

A TEST ITEMS

OUR TEMPERATURE CHART

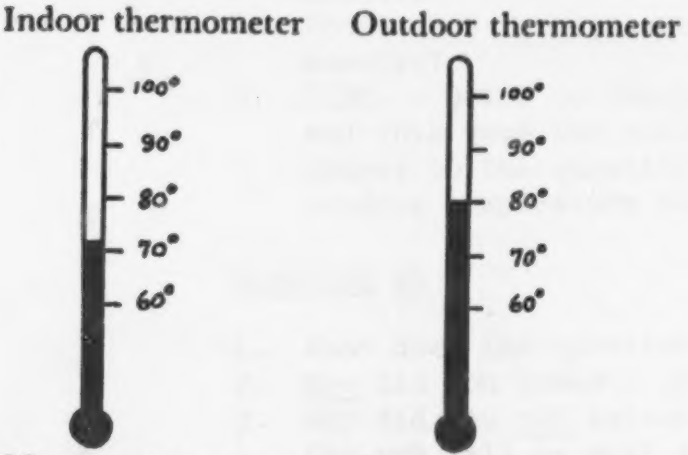
Mary's class kept a record of hourly temperatures, indoors and outdoors, on a spring day. They studied their chart.

Temperature Readings April 1st

	8 a.m.	9 a.m.	10 a.m.	11 a.m.	12 noon	1 p.m.	2 p.m.	3 p.m.
Indoor	62°	71°	80°	75°	72°	74°	75°	72°
Outdoor	60°	67°	71°	78°	80°	78°	77°	75°

20 At what time was the room too cold for comfort?

E 8 a.m.
F 11 a.m.
G 1 p.m.
H 3 p.m.



22 The thermometers drawn here show the temperature as it was at

E 9 a.m.
F 10 a.m.
G 12 noon
H 2 p.m.

24 The time of day when there was the greatest difference between indoor and outdoor temperatures was

E 8 a.m.
F 9 a.m.
G 10 a.m.
H 12 noon

25 The time of day the children most needed to open the classroom windows to make the room comfortable was

A 8 a.m.
B 9 a.m.
C 10 a.m.
D 12 noon

SCIENCE STEP FORM 4B PART II ITEMS 20, 22, 24, 25 (continued)

B. INTERVIEW GUIDE

Non-probing portion

Give problems 20, 22, 24, 25 with responses.

Probing portion

Question 20

1. What does the question ask you to do?
2. What was the temperature indoors at (time give as answer)?
3. What was the temperature outdoors at (time given as answer)?
4. Why did you give the answer you did? (INT. - Be sure to get process - e.g., looked at indoor and outdoor temperatures for 8 a.m., etc.)
5. What is considered the ideal temperature for a classroom? Why? (INT. - Supply ideal temperature if subject does not know).

Question 22

1. What is it that the question asks you to do?
2. How did you arrive at the reason for your answer? (INT. - here, probe for process only..e.g., "What did you do first?")
3. About what temperature is the reading of the indoor thermometer?
4. About what temperature is the reading of the outdoor thermometer?
5. (INT. - point to thermometer charts) If this were the indoor and this were the outdoor thermometers, what would be the right answer to the question? (Interviewer reverses the indoor and outdoor temperature readings).

Question 24

1. What does the question ask you to do?
2. How did you come to your answer?
3. Why did you not select choices (--), (--), or (--)?
4. Can you tell me what the difference in the temperatures was at (response choices)?

Question 25

1. What does the question ask you to do?
2. At what time did they probably open the windows?
3. Why would you not open the windows at 12 noon?
4. Why would you not open the windows at 9 a.m.?
5. Why would you not open the windows at 8 a.m.?
6. Why would you open a window on a cold day?

SCIENCE STEP FORM 4B PART II ITEMS 20, 22, 24, 25 (Continued)

C. SCORING PROCEDURES

1. Possession of information
   All information required to answer this set of questions is given. Rate 3, unless you can find some unusual excess or defect of knowledge.

2. Retrieval of information
   See possession of information above.

3. Extracting information
   - (a) Understands the question (1)
   - (b) Ability to read temperature chart (2, 3)
   - (c) Question 22. Ability to read indoor and outdoor thermometers (3,4)
   - (d) Question 24. Does subtract (e.g., indoor and outdoor thermometers in answering interview questions 3 and 4)
   - (e) Question 25. Utilized temperature chart in answering interview questions 2, 3, 4, and 5.

4. Combining operations

Question 20

Arrives at answer by a process of (a) grasping the meaning of the question; (b) searching the row labeled "indoor" for the lowest temperatures or a temperature below his notion of ideal or to check the temperature given for each alternative E, F, G, H. See interview question 4.

Question 22

Rating	5-4	Adapts logical, orderly, efficient procedure in arriving at answer
	3-2	Some confusion but eventually some semblance of order
	1-	Total lack of a logical, orderly, defensible procedure

Question 24

Understands the question and takes into consideration all needed data supplied in the question in arriving at answer (2, 3)

Question 25

- (a) Evidence of reasoning that windows are opened to change temperature of room in order to make more comfortable (6)
- (b) Evidence of combination with 4(a) of some notion of ideal temperature
- (c) Evaluation of the effects in terms of comfort of opening the window at each hour indicated in an alternative (3, 4, 5)

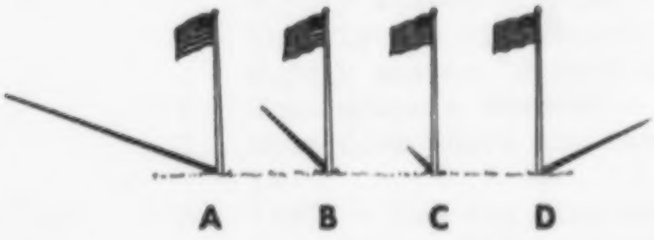
SCIENCE STEP FORM 4B PART I ITEMS 1, 5, 16
(Used for subjects grades 4-6)

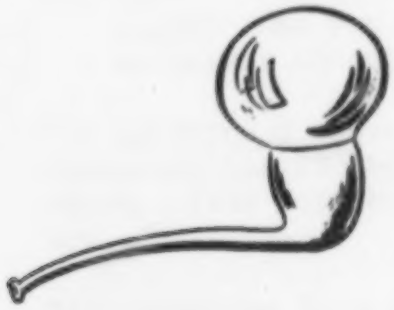
A TEST ITEMS

CAMP SPRUCE

Jack, Bobby, Joe and Harry went to Camp Spruce one summer. At camp, many interesting things happened to them.

1 During the day they noticed the changing shadow of the camp flagpole. Which of the pictures shows the shadow at noontime?



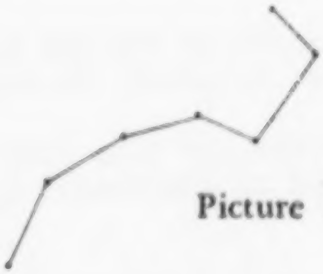


16 Mary bought her little sister, Ann, a bubble pipe to play with. When the bubble was the size of the one in the picture, Ann took the pipe out of her mouth. What do you think happened to the bubble?

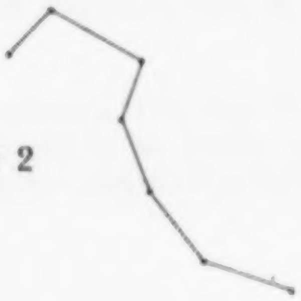
E It burst with a POP.
F It flew up into the air.
G It got smaller and smaller.
H It stayed on the pipe without getting bigger or smaller.

5 Early in the evening Joe found the Big Dipper in the sky. It looked like the dipper in Picture 1. He woke up sometime in the night and noticed it again. It looked like Picture 2. Why did it change position or seem to change positions?

A The stars go around in a circle.
B Stars are always revolving around each other.
C Stars move with the earth.
D The earth turns on its axis.



Picture 1



Picture 2

SCIENCE STEP FORM 4B PART I ITEMS 1, 5, 16 (Continued)

B. INTERVIEW GUIDE

Question 1

Non-probing portion

Present problem and responses

Probing portion

1. Tell me in your own words what the question asks you to do.
2. Why did you answer the way you did?
3. (INT. - Points to shadow of flagpole A) What is this?
4. What is the difference between shadows A and C? (INT. - Supply answer "length" if subject cannot respond.)
5. What is the difference between shadows A and D? (INT. - Supply answer "direction" if subject cannot respond.)
6. What makes a shadow?
7. What does where the sun is have to do with how long the shadow is?
8. (INT. - Use arc diagram) Where would the sun be along this line in order to make a shadow like picture D? Would it be in position 1, 2, 3, 4, or 5?*
9. (INT. - For enactive representation with golf tee and flashlight) Can you use this golf tee and flashlight to help find an answer to question 1?**

*In question 8 above, a card is shown which contains a drawing of flagpoles (as in the test item) with an arc for the sun's movement added. Five possible positions for the sun are depicted along the arc.

**In question 9, a golf tee was used as a substitute for the flagpole. The subject was given a flashlight with which he could produce shadows of the golf tee of different lengths and different directions.

C. SCORING PROCEDURES

Question 1

1. Possession of information
    - (a) Knows what a shadow is (6)
    - (b) Knows that longer shadows are cast when the sun is low in the sky (7)

SCIENCE STEP FORM 4B PART I ITEMS 1, 5, 16 (Continued)

C. SCORING PROCEDURES - question 1 (continued)

2. Retrieval of information
   High rating for retrieval during unstructured portion of interview.

3. Extracting information
   (a) Understands the question (1)
   (b) Recognized shadow in diagram (3)
   (c) Notices differences in length of shadows in diagram (4)
   (d) Notices differences in direction of shadows in diagram (5)

4. Combining operations
   (a) Is able to provide a logical explanation for answer given on interview question number 2. (5-4 rating)
   (b) Is able to answer satisfactorily interview question 8 and 9 even though answer to interview question 2 is inadequate. (3-2 rating)
   (c) Neither of above 4(a) or 4(b) adequate (1 rating)

B. INTERVIEW GUIDE

Question 5

Non-probing portion

Present problem in free-response form, i.e., subject is not shown the four alternate responses.

Probing portion

1. Can you tell me what the question asks you to do?
2. What do the dots in the outline of the dipper stand for?
3. How are these two pictures the same?
4. How are these two pictures different?
5. (INT. - Introduce multiple choice version of question)
6. (INT. - Again ask) What does the question ask you to do?
7. Probe for understanding of meaning of "axis" and "revolution"

C. SCORING PROCEDURES

Question 5

1. Possession of information
   (a) Knows meaning of "axis" and "revolving" (7)
   (b) Knows that it is the movement of the earth on its axis and not the movement of the stars that result in the apparent difference of the dipper at different times of the night.

SCIENCE STEP FORM 4B PART I ITEMS 1, 5, 16

C. SCORING PROCEDURES (Continued)

2. Retrieval of information
   Evidences recall of 1(a) and 1(b) in open ended form of question

3. Extracting information
   (a) Understands the question (1)
   (b) Is able to tell what the dots stand for in the diagram (2)
   (c) Is able to tell how the two pictures of a dipper in the diagram are the same and how they are different (3, 4)

4. Combining operations
   Rating 5-4 Uses understanding of the movement of the earth on its axis in selecting D or elimination A, B and C
   3-2 Uses logical process in eliminating at least one alternative (or selecting one alternative)
   1- Fails to combine provided and stored information in a logical fashion

B. INTERVIEW GUIDE

Question 16

Non-probing portion

Present problem in free response form.

Probing portion

1. What does the question ask you to do?
2. Why did you give the answer you did? (INT. - Supply correct answer if subject answer is incorrect)
3. What is a bubble?
4. How can you make a bubble? (At least 3 ways)
5. Why does a bubble get smaller?
6. What makes a bubble bigger?
7. What has happened when a bubble bursts? (INT. - Probe until it is evident the answer is based on more than an experimental level.)
8. Which of the following works the same way as the bubble? For each give me a "yes" or "no" answer
   (a) A bicycle tire puncture
   (b) A rubber ball bouncing
   (c) Turning a water tap on
   (d) Sharpening a pencil

SCIENCE STEP FORM 4B PART I ITEMS 1, 5, 16 (Continued)

C. SCORING PROCEDURES - Question 16

1. Possession of information
   Is able to answer adequately interview questions:
   What is a bubble? (3)
   How can you make a bubble? (4)
   Why does a bubble get smaller and smaller? (5)
   What makes a bubble get larger? (6)
   What has happened when a bubble bursts? (7)

2. Retrieval of information
   Evidences recall of science information and/or life experiences with bubbles.

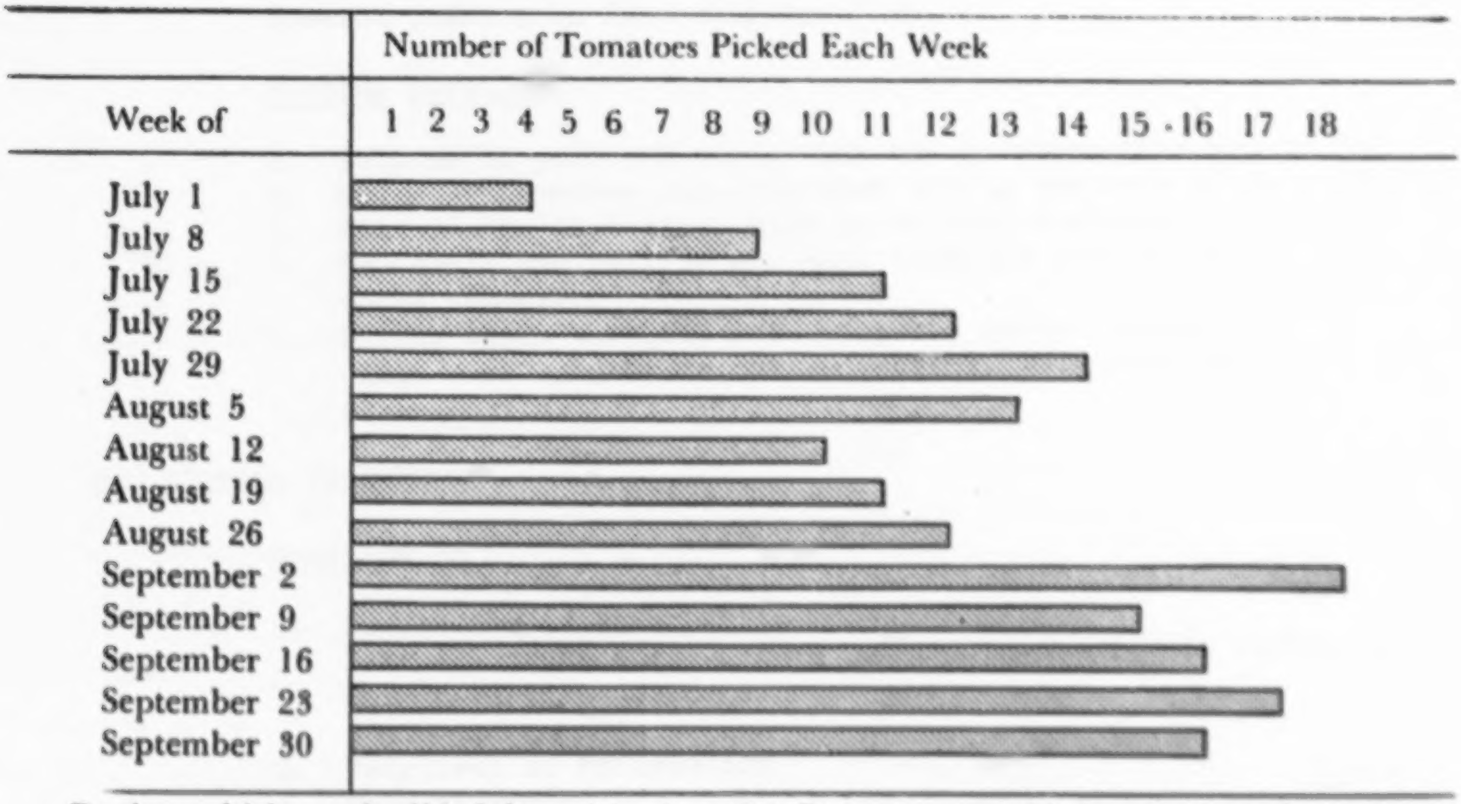
3. Extracting information
   (a) Understands question

4. Combining operations
   (a) Applies science information to question
   (b) Is able to answer correctly interview question 8
   (c) Applies life experience information to question

SCIENCE STEP FORM 4B PART I ITEMS 22 and 24
(Used for subjects grades 4-6)

A. TEST ITEMS

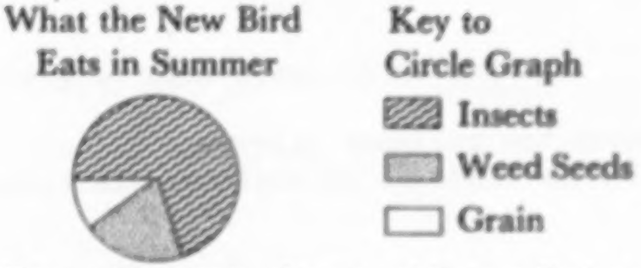
22 John counted the tomatoes picked from his garden each week. He kept this record.



During which week did John pick the most tomatoes?

E July 22
F August 5
G September 2
H September 23

24 Sue saw a new bird come to their garden. In a book she found its picture and a circle graph to show what the bird eats in summer.



From this graph, she was able to tell that this bird was

E helpful and ate mostly insects
F helpful and ate grain
G helpful and ate mostly weed seeds
H harmful and ate mostly grain

SCIENCE STEP FORM 4B PART I ITEMS 22 and 24 (Continued)

B. INTERVIEW GUIDE

Question 22

Non-probing portion

Present problem in free-response form

Probing portion

1. Tell me in your own words what the question asks you to do.
2. How many tomatoes did John pick during the week of July 29th?
3. What do the dark lines have to do with tomatoes?
4. How could you tell me how many tomatoes John picked in the whole summer?
5. During which week did John pick the fewest tomatoes? (INT. - If subject is not successful, give four responses and probe further)

C. SCORING PROCEDURES

Question 22

1. Possession of information
   Rate all protocols 3 unless unusual possession or defect in information is noted.

2. Retrieval of information
   Same as above

3. Extracting information
   (a) Understands the question (1)
   (b) Demonstrates ability to read graph (2, 3)

4. Combining operations
   (a) Is able to apply ability to read graph in answering question 4 and 5
   (b) Searches for longest line in answering question and notes week, and then evaluates alternatives to multiple-choice question.

B. INTERVIEW GUIDE

Question 24

Non-probing portion

Present problem with responses.

SCIENCE STEP FORM 4B PART I ITEMS 22 and 24

B. INTERVIEW GUIDE (Continued)

Probing portion

1. What does the question ask you to do?
2. How did you select your answer?
3. How can a bird be helpful?
4. How can a bird be harmful?
5. How can a bird be helpful if he eats mostly insects?
6. How can a bird be harmful if he eats mostly grain?
7. Show me the part of the circle that stands for insects?
8. What does the bird eat most of?
9. How are weed seeds harmful?
10. How are insects harmful?

C. SCORING PROCEDURES

Question 24

1. Possession of information
    (a) Knows some ways that birds are helpful and some ways birds are harmful (3, 4)
    (b) Knows why it is helpful for a bird to eat insects (5)
    (c) Knows why it is harmful for a bird to eat grain (6)
    (d) Knows how weed seeds are harmful
    (e) Knows how insects are harmful

2. Retrieval of information
    Evidences some retrieval of the above knowledge in selecting answer (2)

3. Extracting information
    (a) Understands the question asked (1)
    (b) Is able to show part of the circle that stands for insects (5)
    (c) Is able to read circle graph (8)

4. Combining operations
    (a) Combines understanding of question, ability to read graph, and retrieval of information in arriving at answer (best source in rating - question 2)

SCIENCE STEP FORM 3B PART I ITEM 7
(Used for subjects grades 7-9)

A. TEST ITEM

7 Harry did the following experiments to learn something about what causes house paint to dry

1. He painted a board and placed it in the sunlight.
2. He painted a second board and placed it where it could get no sunlight.
3. He painted the third board and pressed a sheet of wax paper onto the paint.
4. He painted a fourth board and placed it in a bucket of water.

The next day he found that only the paint on the first and second boards had dried. This was because

A. air is necessary for the drying of paint
B. air and sun are necessary for the drying of paint
C. air is not necessary for the drying of paint
D. sunlight is necessary for the drying of paint

B. INTERVIEW GUIDE

Non-probing portion

Present problem and responses

Probing portion

1. What does the problem ask you to do? (Be sure to find out whether the subject understands that he is to select response A, B, C, or D on the basis of the information contained in 1, 2, 3, and 4 and the results of the four experiments.)
2. What was your reasoning that led you to select (--)?
3. What do experiments 1 and 2 have in common?
4. What do experiments 3 and 4 have in common?
5. (If subject fails to state that in 1 and 2 the paint dried and in 3 and 4 it did not, ask:) In which of the four experiments did the paint dry?
6. What does sunshine have to do with drying paint? Why?
7. What seems to be the important thing that Harry did with the first and second board that he did not do with the third and fourth board? Why do you think the paint on the first and second boards dried?
8. Why is B not a good answer to this question?
9. Why is C not a good answer to this question?
10. Why is D not a good answer to this question?
11. How could Harry have made his experiment better?
12. (Ask only if subject did not give a good answer to 11) Would it have made any difference if Harry had used different kinds of paint on different boards so long as the paint was all the same color?

SCIENCE STEP FORM 3B PART I ITEM 7 (Continued)

C. SCORING PROCEDURES

1. Possession of information
   All the information required to solve the problem is presented in the test item.

2. Retrieval of information
   Rate 3 unless the unusual recall or a failure to recall facilitates or hinders the solution of the problem.

3. Extracting information
   (a) Understands the question (1)
   (b) Notes that experiment 1 and 2 the paint dried (3)
   (c) Notes that in neither experiment 3 nor 4 the board did not dry (4)
   Equally good answers for (b) and (c) involve combining operations. In experiment 1 and 2 the boards are exposed to air; in experiments 3 and 4, they are not.
   (d) Experiments 1 and 2 had air, experiments 3 and 4 did not

4. Combining operations
   Uses experimental evidence to eliminate B and C and D and select A.

SCIENCE STEP FORM 3B PART I ITEM 20
(Used subjects grades 7-9)

A. TEST ITEM

20 Tom was told that he should plant radishes in the dark of the moon or they would all "go to tops" and the size of the radishes would be quite small. Which one of the following could he do to determine the truth of the statement?

E Tabulate the opinions of ten experienced gardeners on the effect of the phase of the moon on plant growth.

F Plant some radish seeds at different phases of the moon and compare the size of the radishes when each was four weeks old.

G In a gardening book, read about radishes and how to plant them.

H Inspect carefully radishes that are growing in several different gardens.

B. INTERVIEW GUIDE

Non-probing portion

Present problem and responses

Probing portion

1. What does the question ask you to do?
2. Why did you select the answer you did?
3. What does the "dark of the moon" meen?
4. What does "go to tops" mean?
5. What does "tabulate" mean? (Supply information to 3, 4, 5, if lacking)
6. (If subject was initially unsuccessful) ask: Would you please answer the question again as if you were a scientist?
7. Why would you not select (--), (--), or (--)?

C. SCORING PROCEDURES

1. Possession of information
   - (a) Knows meaning of "dark of the moon" (3)
   - (b) Knows meaning of "go to tops" (4)
   - (c) Knows meaning of "tabulate" (5)

2. Retrieval of information
   There is no problem of retrieval here as the terms to be understood are in the question.
   Rate 3

SCIENCE STEP FORM 3B PART I ITEM 20

C. SCORING PROCEDURES (Continued)

3. Extracting information
   (a) Understands the question (1)

4. Combining operations

Rating 5- (a) Assumes an experimenter's set
(b) Reasons to answer
(c) Provides defensible reasons for elimination of alternatives (7)

4- Assumes set after question 6

3-2 (b) Reasons to answer
(c) Provides defensible reasons for eliminating alternatives not chosen (7)

1- None of the above

SCIENCE STEP FORM 3B PART II ITEMS 2 and 3
(Used for subjects grades 7-9)

A. TEST ITEMS

A general science class was engaged in the study of a unit of foods. Several small groups had been formed to make studies of food problems. In their studies they used the table shown below.

PERCENTAGE COMPOSITION (BY WEIGHT) OF COMMON FOODS

Food	Carbohydrates	Fats	Proteins	Minerals	Water
Apples	14.9	0.4	0.3	0.3	84.1
Bacon, lean	1.1	53.0	12.2	4.7	29.0
Beef, round	----	13.0	19.1	0.9	67.0
Bread, whole-grain	47.0	4.0	10.0	2.0	37.0
Cake, frosted	62.1	6.2	5.2	1.3	25.2
Candy, Chocolate bar	57.7	33.5	6.0	1.7	1.1
Crackers, soda	72.7	9.6	9.6	2.4	5.7
Dates	75.4	0.6	2.2	1.8	20.0
Egg	0.7	11.5	12.8	1.0	74.0
Ham, smoked	0.7	25.0	19.5	5.8	49.0
Jellies	65.0	----	0.2	0.3	34.5
Milk, whole	4.9	3.9	3.5	0.7	87.0
Peanut Butter	21.0	47.8	26.1	3.4	1.7
Potatoes, white	19.1	0.1	2.0	1.0	77.8

2 The same group reported that carbohydrates are very important in the diet. Carbohydrates and fats furnish energy. The food which supplies the highest total percentage of fat and carbohydrates is
E frosted cake
F chocolate candy bar
G whole milk
H soda crackers

3 Another group reported on food preservation. They told the class that by using the table they had figured out how much 100 pounds of raw eggs would weigh if all the water were taken out of them. Which of the following answers did they obtain?
A 20 pounds
B 26 pounds
C 32 pounds
D 36 pounds

B. INTERVIEW GUIDE

Non-probing portion

Give question 2 and 3 before probing for question 2.

SCIENCE STEP FORM 3B PART II ITEMS 2 and 3

B. INTERVIEW GUIDE (Continued)

Probing portion

Question 2

1. What does the question ask you to do?
2. What is the meaning of "percentage"?
3. What is the meaning of "diet"?
4. What food has the highest percentage or protein?
5. How did you figure out your answer to question 2? (INT. - Here we want to know the sequence of steps used by the subject)
6. How can you know that (--) is correct?
7. Why are (--), (--), and (--) not correct?
8. If you had four ounces of cake and two ounces of candy bar, would this change your answer?

Question 3

9. What does the question ask you to do?
10. What percentage of smoked ham is water?
11. How did you figure out your answer? (INT. - Get sequence of steps)
12. How can you know that (--) is correct?
13. Why is (--), (--), or (--) not correct?
14. Card Sorts: "What cards help you to solve the problem?"
    (a) eggs are mostly water
    (b) eggs are made up of 74% water and 26% non-water
    (c) eggs are 11.5% fats

    (a) 100 pounds of eggs is 74% water
    (b) eggs weigh less than 1 pound each
    (c) eggs contain less than 1% carbohydrates

C. SCORING PROCEDURES

Question 2

1. Possession of information
    (a) Knows meaning of "percentage" (2)
    (b) Knows meaning of "diet" (3)

2. Retrieval of information
    Rate 3

3. Extracting information
    (a) Understands the question (1)
    (b) Ability to get information from chart (4)
    (c) Notes figures are percentages and the question asks for the highest percentage (8)

SCIENCE STEP FORM 3B PART II ITEMS 2 and 3

C. SCORING PROCEDURES (Continued)

4. Combining operations
    (a) Logical sequence of steps used to arrive at answer
    (b) Logical reason for selection of answers or (verification of hypothesis)
    (c) Logical process for elimination (or rejection of hypothesis)

Question 3

1. Possession of information
    (a) Understands percentage concept (10)

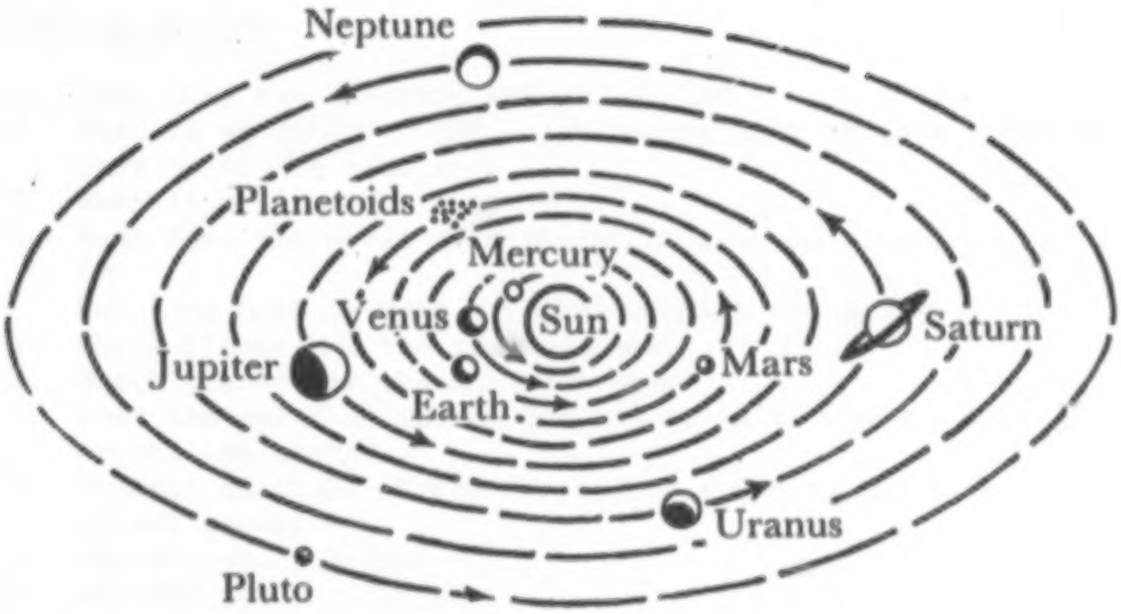
2. Retrieval of information
    Rate 3

3. Extracting information
    (a) Understands the question (9)
    (b) Able to read table (10, 11)
    (c) Selects relevant information (card sort)

4. Combining operations
    (a) A defensible sequence - example: Looked for percentage of water in eggs; computed number water column on egg row; noted 74% of eggs are water; therefore 26% non-water; 26% of 100 pounds = 26 pounds; notes answer B = 26 pounds
    (b) Logical process (11, 12)

SCIENCE STEP FORM 3B PART II ITEM 28

A. TEST ITEM



28 Henry looked at the diagram shown above in a booklet his leader gave him.

If the planets all travel at the same speed in their orbits, which planets have a shorter year than our year?

E Mercury and Venus
F Mars and Venus
G Pluto and Jupiter
H Mercury and Mars

SCIENCE STEP FORM 3B PART II ITEM 28
(Used for subjects grades 7-9)

B. INTERVIEW GUIDE

Non-probing portion

Present problem in free-response form

Probing portion

1. What does the question ask you to do?
2. What is an orbit? (INT. - If subject has trouble verbalizing: Show me on the picture)
3. What is a planet?
4. What does the orbit of a planet have to do with a year?
   OR
   How long does it take a planet to make one orbit?
5. Which planet has the longest year?
6. Which planet has the shortest year?
7. Does the earth go around the sun or is it the sun that goes around the earth?
8. Present multiple-choice version of question and have subject select answer.
9. Why did you select (--)?
10. Why did you not select (--), (--), or (--)?

C. SCORING PROCEDURES

1. Possession of information
   (a) Knows meaning of orbit and planet (2, 3)
   (b) Knows what the orbit of a planet has to do with a year OR how long it takes one planet to make one orbit
   (c) Knows that the earth goes around the sun; not the sun going around the earth (7)

2. Retrieval of information
   Rating 5- Evidences some understanding of 1(a), (b), (c) in structured portion of interview
   4- Evidences understanding of 1(a), (b), (c) in answering interview questions 9 and 10
   3-2 Some evidence of lack of retrieval
   1- Clear evidence of lack of retrieval

3. Extracting information
   (a) Understands the question
   (b) Focuses attention on "planets travel at the same speed"
   (c) Notes distance of planets from the sun

SCIENCE STEP FORM 3B PART II ITEM 28

C. SCORING PROCEDURES (Continued)

4. Combining operations
   Uses knowledge that planets travel at the same speed and one orbit takes one year to deduce that shorter orbits will mean shorter years.